Vol. 42, No. 4 April, 1945

Psychological Bulletin

A SYSTEMATIC TREATMENT OF JUDGMENT

DONALD M. JOHNSON
University of Illinois

The concept of judgment has had a long and interesting history in psychology and logic. The variety of meanings carried by this term is disclosed by Hollingworth's (50) review of the traditional definitions of judgment and by Weber's (111) later questionnaire study of its usage by psychologists and logicians. The topic of this paper can best be delimited, however, by making explicit the usage of modern experimental psychology.

In many situations—in the laboratory and out—a person is called upon to decide between two or more alternatives or categories of response. The alternatives may be in the perceived situation or they may be reproduced from memory. Or, as in complex problem situations, the alternatives may be hypotheses or tentative solutions which result from a prolonged search, and judgment of the alternatives in these cases comes in as one phase of the larger enterprise of problem solving.* As differentiated from other higher mental processes judgment is decisive, not productive. It is decisive in a functional sense in that it closes an episode of deliberation and permits the resumption of other kinds of activity. It is not productive since nothing new is added to the perceived situation as by imagination, memory or generalization. Treating judgment in this way is, of course, a device of systematization which skips over some of the niceties of traditional usage. It does accord with the practice of experimental psychology and with the usage of at least one respectable logician, for Dewey (27) speaks of judgment throughout his recent book as "settlement of a problematic situation."

Approaching the topic with this orientation this paper attempts to organize the literature on judgment, describing some of the methods in present use, the assumptions commonly made, and summarizing in general terms some of the achievements already at hand. The first section takes up the independent variables of judgment, the materials, factors or stimulus objects on which the judgment is based. Following this is a section on the dependent variables, categories of response, confidence

* In a previous review in this *Bulletin* (62) the elaboration of solutions to a problem was discussed at some length, but judgment of alternative solutions was left for this paper.

and time of judgment. Finally, individual differences in these various phases of the topic are brought together.

I. Independent Variables of Judgment

Some important conditions of judgment, such as attention and motivation, will not be discussed here. Attention is obviously important in any intellectual activity; one cannot judge an issue if he cannot attend to it. But the role of attention in problem solving has been considered in the previous paper, and the same considerations apply to judgment. The motivation of an act of judgment—as it is treated in the literature—seems to be similar to the motivation of any other activity, and the activity continues until the motivation is reduced, whether by a decision which clarifies the situation or by some less logical change. Blumenfeld (9, pp. 168–173) has a discussion of this problem in which he considers a variety of occasions for judgment. He is not able to come to any very specific conclusions, however. The independent variables of judgment which have been investigated most thoroughly are the stimulus materials to which the response is related. Studies on this phase of judgment can easily be divided into two groups: those dealing with simple judgments and those dealing with complex judgments.

A. Simple Judgments

In many laboratory experiments the objects of judgment vary only (or mainly) along a single dimension or continuum, and it is this aspect of the objects which is judged. It is relatively easy in these judgments to isolate and attend to the aspect or quality designated in the instructions. Simple judgments can be grouped for later reference into three large classes on the basis of the aspect or characteristic of the stimulus objects which is to be judged.

Some of the most valuable data on judgment come from *perceptual* judgments. In these the observer attends to and judges some perceived aspect of the stimulus objects: color, loudness, breadth, distance, etc. The rationale of the psychophysical methods is, in large part, an attempt to control the peculiarities of the judging process, or to balance them out experimentally or statistically, so that the results can be clearly related to the perceptual mechanisms. It is precisely these peculiarities of judgment, e.g., constant errors, which concern us in this paper; psychophysical research will be mentioned only when it illuminates the judging process in general. And since under some conditions perceptual judgment becomes a matter of simple perceptual discrimination, it is necessary to outline a theory of discrimination developed by Landahl, one of Rashevsky's students at Chicago.

Landahl (70) begins with Rashevsky's two-factor theory of neural excitation which states that the excitation factor increases at a rate proportional to the

excitation and decays at a rate proportional to its own concentration. He further assumes that the excitation is a logarithmic function of the intensity of the stimulus divided by the stimulus threshold. Then he sets up a model beginning with two sensory pathways and ending with two response pathways. Between these are four connecting pathways making synaptic junction, one excitatory and one inhibitory to each. By making simple assumptions Landahl can write an expression for the preponderance of excitation in one response pathway over the other as a function of relative stimulus intensities.

He then introduces the concept of "random distribution of distracting stimuli" at a synapse. Working with a symmetrical distribution of these distracting stimuli, he comes out with the probability, for certain cases, of responses such as *greater*, *less* and *equals*. Thus he can take two values from a weight-lifting experiment published by Guilford and compute the other values with fair accuracy.

In a later paper (**71**) Landahl extends the theory to complex stimuli—including non-physical stimuli—and thus is led to the possibility of factor analysis for separating factors producing the total excitation at the two pathways. These deductions have been based on the assumption that the stimuli are operating simultaneously. To take account of delayed discrimination Landahl (**72**) uses a more complex model of neural pathways which includes a set of self-exciting neural circuits of low thresholds.

Householder (**51**, **52**), also a Rashevsky student, attempts a similar problem, namely, to describe—within the framework of mathematical biophysics—a mechanism by which differences in intensity of stimulation are sifted into different sets of neural pathways and thus produce different responses.

The chief contribution of these papers is, of course, methodological. The authors are more interested in exploring the potentialities of mathematical biophysics than in settling any particular problem. A diligent reader of these mathematical essays will be intrigued by the possibilities of ignoring the details and speculating on the grand scale, and will perhaps find his attention directed to concepts which are more fundamental than those usually attacked in psychophysical experimentation. The idea of a "random distribution of distracting stimuli," to mention one, will come up again in these pages by one name or another. As electrophysiological research advances, it is likely that threshold phenomena, receptor functions, and the like, which Landahl and Householder are forced to guess at, will be analyzed experimentally, and that theoretical speculation and empirical data can draw closer together. Already some data on the relation between electrical activity in the muscles and judgments of lifted weights are available (**40**, **88**).

Experiments on *affective* judgments have contributed a large share of present knowledge of the psychology of judgment. In these the observer reports on his likes and dislikes, how pleasant or unpleasant the stimulus object is to him, how badly he wants something, how strongly he reacts for or against something. The older experiments in this field used colors, musical intervals, odors and forms. More recently faces, words, activities, slogans and the like have been included. The term "affective" is used also in a narrow, more precise sense; in the broad sense used here it

comes close to "personal," as opposed to impersonal or objective. The variety of materials on which such judgments have been made may be seen in Young's (116, pp. 270–292) recent summary. Such judgments are easily made; in fact we shall see that they are often made when other kinds of judgment are called for.

A third class of judgments, *abstract* judgments, may be distinguished, in which attention is directed to abstract or conceptual properties of the stimulus objects, and the judgments are made on this basis. In a library, for example, the cataloger who has to decide whether to list a new title, say *Growth of American Thought*, under history, philosophy or genetic psychology is making such an abstract judgment. The judge in a court of law who rules on the admissibility of evidence supplies another illustration. In studies of personality judges are often asked to rate people on such abstract qualities as social usefulness and honesty and to rate occupations on a scale of prestige. These judgments are sometimes difficult and, if we may go by the research on syllogistic judgments as summarized by Woodworth (115, pp. 810–817), a large share of the difficulty is due to the very abstractness of the material.

Refined quantitative research on abstract judgment is meager. As with other kinds of judgment the first problem is one of stimulus control. Our understanding of perceptual mechanisms has been facilitated by the availability of a wide range of stimuli, accurately controlled and suitably spaced along the stimulus continuum. Recent developments in the scaling of attitude items have two important implications for the psychology of judgment. Firstly, they show that the judges who aid in the construction of a Thurstone-type attitude scale—educated adults with an interest in the task—are able to conceive an abstract stimulus continuum, e.g. favorableness of the item toward Communism, to compare items in this respect, and even to sort the items directly in accordance with their location on the abstract continuum. They are able to attend to this abstract aspect of the item in spite of interference from other aspects, foreign to the attitude continuum, and even from the personal acceptability of the item (32, 48, 89). Furthermore, the scale values derived from abstract judgments made in various ways are rather closely related (33, 93), a fact which rules out the possibility that these scale values are artifacts. The judges are judging some genuine abstract content of the items. Secondly, these scaling procedures solve the problem of stimulus control for abstract judgments. They make possible the manufacture of stimulus material to definite specifications in regard to ambiguity and location on a conceptual or ideational continuum. While these scaling procedures have had their chief application in the construction of attitude scales, there seems to be no good reason, *a priori*, why any abstract quality at all, however subtle, cannot be manipulated in the same way. A minor limitation is that the scale values are derived

from the consensus of a group and may not be precisely valid for any one individual.

Abstract material prepared in this way has been used in a few experiments on judgment. Chant and Salter (21) used Droba's scale of attitude toward war to show the relation between difficulty of a judgment and the galvanic skin reflex. Johnson (60) used such material to study the relation between confidence in a judgment and distance from the category threshold. Cartwright (18) had his subjects scale names in respect to radicalism-conservatism in order to get stimulus material for a study of decision time. McGarvey (80) had her subjects scale occupations in respect to prestige and certain forms of behavior in respect to undesirability in order to study the anchoring of the scales. Mosier (84) has scaled over 200 adjectives from the Thorndike Word List on a scale of favorableness-unfavorableness. The results of these studies will be considered later in a systematic way.

The question which logically arises at this point is: What is the mechanism by which these abstract judgments are made? In this form the question has not been attacked, but it is part of the general problem of the psychology of learning, and some of the recent work on generalization and differentiation, on stimulus equivalence and the like shows the path where the answer may be found. The method used in studying the generalization of a scale of values (63) also offers suggestsion.

B. Complex Judgments

Under this heading come investigations in which the objects of judgment vary in many ways, and the judgments are made in respect to some rather general criterion. Such judgments occur in many practical situations—more frequently of course than the simple judgments—as when teachers grade students' schoolwork, when inspectors sort mechanical products, when farmers judge corn at the county fair, when an employer sizes up a job-seeker, when the voters pick a public official, and when a judge or jury decides the fate of a defendant. Studies of the independent variables involved in these complex judgments can be reviewed under two subheadings: identification and weighting of the variables, and interaction among the variables.

Identification and Weighting. Much of experimental psychology is devoted to an analysis of the independent variables or factors which determine a specific response, but our concern here is with general principles of judgment, even though tentative, which can be extracted from the experimental literature.

One such principle, almost obvious, is that *the judge may not be able to point to the factors by which his judgments are determined.* A corollary to this is that a psychologist designing an experiment may not be able to predict his subjects' responses on the basis of logical relationships between the independent variables and the response. This principle is not at all new, and no sophisticated psychologist expects people to be completely rational. It is mentioned here only for the

sake of completeness, since the principle is enthusiastically rediscovered from time to time.

A second principle is that *the judgment may be determined by an independent variable which the subject is not aware of even when his attention is directed to it.* It is now well established that, under some conditions, people can respond with an accuracy above chance levels to subliminal stimuli, i.e. to stimuli which have been reduced in intensity below the conscious threshold (**5, 82, 113**). Furthermore, Bressler's (**11**) experiment on the debated question of the effectiveness of the Müller-Lyer illusion when the radiating lines are subliminal seems to prove that these accessory subliminal stimuli influence the judgment of the main object of attention. It has been suggested (**85**, p. 163) that our immediate intuitive judgments of personality are based in part on such subliminal cues, subliminal in a perceptual sense or in the sense that they depend on memory impressions too faint to reach the conscious level.

Our third generalization must be stated tentatively: *if the judgment called for is difficult, judgment in terms of some other criterion is likely.* Thornton (**104**) showed that, when college students are shown a photograph and asked to rate the person as to various personality traits—an almost impossible task—a person smiling is likely to be rated higher in honesty than the same person not smiling, and a person wearing glasses is likely to be rated higher in intelligence, dependability, industriousness and honesty than the same person not wearing glasses. In a further study Thornton (**105**) found that the effect of the glasses was less when the actual people were seen than when slides of these people were shown. As the author suggests, these factors "will have decreasing effect upon judgments of personality traits as the number of other cues upon which judgments may be made increases." In line with this principle is the limitation on the effects of suggestion (**22, 74**) and on the effects of one's desires (**16, 81**) when the situation is well structured, i.e. when judgment is easy.

Our fourth generalization, which could be considered a special case of the third, is that *when an abstract judgment is called for, an affective judgment is commonly given.* This phenomenon is regularly observed in judgment of the controversial issues of the day, since these are abstract and difficult, and the content is usually affective. Laboratory evidence is also available (**16, 77, 81**). In his original study Lund (**77**) attempted to weight the importance of desire and of evidence in the determination of the judgment, and he reports a correlation of .81 between degree of desire and degree of belief, to be compared with a correlation of .64 between degree of knowledge and degree of belief. "Degree of knowledge" here means merely how often the group stated that their belief was based on knowledge rather than opinion, hence these correlations cannot be taken as indicating the relative weight of these two variables in the judgment. Bird (**8**) has written a critical discussion of this kind of research.

A fifth principle relates to the source of the material. *In general people are likely to weight positively material which comes from, or is attributed to, someone whom they respect* (**4, 76, 92**). Lurie (**78**) reverses the procedure and uses this effect as a measure of prestige. The relative weight of majority and expert opinion in determining judgment of controversial issues has been studied under various conditions (**14, 69, 79, 83**).

Probably other generalizations could be extracted from the experimental literature. Social psychologists have compared the relative weights of emotional and rational material and of various modes of presentation of the material. Of more fundamental significance for the present review would be an experimental study of primacy and recency.

Would a given factor weigh more heavily in the final decision if it is introduced early in the course of deliberation or late?

The methodology of these studies is important. It is comparatively easy by several methods to determine that such and such a factor has some bearing on the judgment. To determine in any general sense the relative weight of several factors is much more difficult. For example, to compare judgments made after exposure to emotional and to rational material, or judgments based on knowledge and on desire, it is necessary to assume that each kind of material used in the experiment is an equally potent representative of its class, i.e. that the rational material is as rational as the emotional material is emotional. If this assumption is not justified, the conclusions may still be useful in a limited way but must be greatly restricted in generality. It is now possible by the scaling procedures mentioned above to make a more precise approach to these problems. If the stimulus material can be located on a scale of rationality, or desirability, or authoritativeness, refined comparisons or correlations can be made.

The correlational technique seems to offer a means to weightings of some degree of generality. It has been used to analyze both perceptual and abstract judgments.

Brunswik (13) has been using the correlational technique on the old psychophysical problem of size constancy. He used a representative sample of objects encountered by one subject as she went about her daily routine, and had the subject make judgments about these objects under several different instructions. If we take as our problem the determination of the judgments of the size of the objects—which was not Brunswik's main concern—we can get some good information from his correlation coefficients. From his Table 5 it appears that the net relationship between estimates of size and actual or "geographic" size was .98, while that between estimates and size of the retinal projections was .40. Brunswik goes on to show, on the basis of his "ecological" sample, that the perceptual system is better adapted for judging distal stimuli (objects in the environment) than proximal stimuli (retinal objects) which mediate the distal judgments.

Osgood and Stagner (87) used the correlational method for analysis of judgments of prestige. They obtained ratings for the prestige of fifteen common occupations, and also ratings for hopefulness, honesty, idealism, hours of work and the like for the same occupations. They then correlated the ratings for these qualities with the ratings for prestige and found that the ratings for hopefulness were high, showing a correlation of .99 with prestige, while ratings for hours of work were low, with a correlation of .20. A similar analysis of the prestige of the people in these occupations was also carried out. Taking these correlations as evidence of the specific factors on which the general judgment was based it may be concluded that "qualities like brains, leadership and self-assurance are shown as marks of prestige" while "honesty, idealism and congeniality are not."

The ideal of this type of analysis—though no one has carried it that far—appears to be a multiple regression equation, identifying and weighting all the variables which have significant effects on the judg-

ment. The limitations of the correlation technique are well known today, and even in 1918 Thorndike (101) in a paper called *Fundamental Theorems in Judging Men* was calling attention to non-linear relations between the independent variables and the judgment. The competent impressionistic or intuitive judge of men, he said, takes these complexties into account, hence the strength of the intuitive judgment, as compared with the more formal types, is that it is more quantitative! In spite of these limitations the correlation technique has a definite value if it leads to a good first approximation and narrows the field of subse-

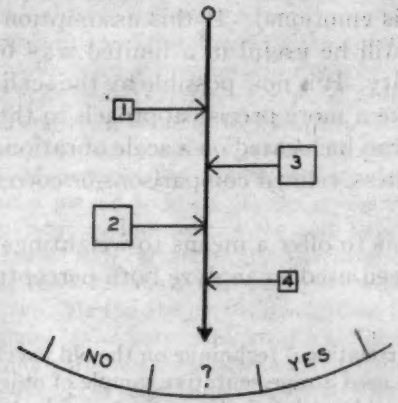

FIG. 1. Illustrating some of the possible modes of interaction of the independent variables which determine a judgment. The numbered arrows indicate stimulus variables, forces, factors, etc. The scale at the bottom is made up of the categories of response. This figure specifically represents the assumption that the response is a result of the algebraic summation of the various contributing factors.

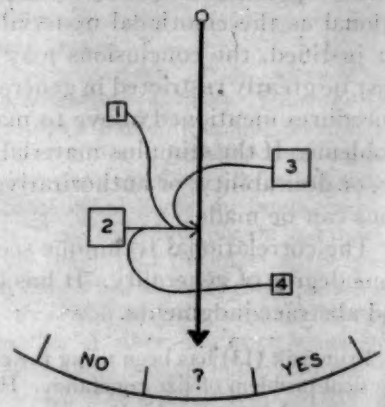

FIG. 2. Judgment on the basis of general impression. Variable No. 2 sets up the general impression and the other variables reinforce, or at least do not oppose this impression. This is illustrated by the *halo effect* and the *atmosphere effect.*

quent inquiry. A more detailed analysis waits for an understanding of the interaction which takes place among the independent variables.

Modes of Interaction. If we knew what variables were involved in a complex judgment, and their approximate weights, we would next like to know something of the dynamics of the process: How do these variables cooperate in producing the judgment? The most obvious mode of interaction, one that is often assumed implicitly, is *summation.* It is possible that, under some conditions at least, all factors which influence the judgment exert their influence in the same way, and the judgment is, therefore, the algebraic resultant of these separate pushes and pulls. Figure 1 attempts to illustrate this mode of interaction.

Since psychometric frequency functions are similar to the integral of the normal probability function, it is often assumed, as by Boring (10), that simple psychophysical judgment depends upon the additive interaction, presumably at the neural level, of a multitude of independent varying "dispositions." Landahl (70) makes this assumption also when he adds the excitation from the sensory apparatus to the excitation produced by the "random distracting stimuli." Whether this assumption or summation at the neural level is justified or not, we find similar assumptions at the psychological level. Mosier (84), for example, in introducing his "psychometric study of meaning," writes a hypothetical equation for meaning (expressed by judgments of words in terms of favorableness-unfavorableness) as the sum of "that part of the meaning which may be taken as constant from one person and one context to another," "the part which varies from one individual to another" and "the portion of the meaning which is due to the context." Cartwright and Festinger (20) in their topological theory of decision likewise assume simple algebraic summation of opposing psychological forces. Max Eastman (28) in his thoughtful discussion of humor makes a slightly different assumption: "The two kinds of drollery, the pun and the mischievous fancy, fuse into a whole that is more comical than the sum of the comicalities of each" (p. 111). "Remember that two humor sources welded together yield more humor than the sum of their two separate yields" (p. 314).

The only critical test of this assumption, as far as the writer can discover, consists of a few studies of affective judgments. The method has been to get ratings of two stimulus objects, e.g. color cards, separately and in combination and to correlate these sets of ratings. Beebe-Center (7) has summarized the evidence, from which it appears that this method is not strictly adequate because the result of juxtaposing the separate stimuli may be, not a combination, but a new configuration. When the separate stimuli are "heterogeneous," as when the object is judged separately for form and for color, the method is more adequate and leads to the conclusion that "the hedonic tone of the composite phenomenon varies more or less directly with the hedonic tone corresponding to the single stimuli" (p. 122). This is not summation in any precise mathematical meaning; one cannot compute the rating of the composite by adding the ratings of the components. Greater refinement of the rating scale or some other way of estimating the strength of the contributing factors is necessary before any satisfactory answer to this question can be obtained.

One can argue that summation is likely under some conditions and that it may be assumed as a first approximation for a mathematical theory—of simple judgments at least. In addition one must recognize the occurrence of other modes of interaction. Perhaps as a reaction to the implicit assumption of summation there has been an emphasis in the literature recently on the *structure* of the situation which is judged. This term, or something similar, is a convenient one for interpreting the formation and shifting of attitudes, but its usage is usually negative in the sense that it is an admission of failure of the assumption that the attitude can be interpreted as the summation of the various pros and cons

into which the issue can be analyzed. Another difficulty with the emphasis on structure is that it may be putting the cart before the horse. Since the judgment, once arrived at, restructures the subject's beliefs and actions, he could readily report—and be himself convinced—that the restructuring determined his judgment. In spite of these criticisms there are at least two kinds of structuring, or non-additive modes of interaction, which can be described with some definiteness.

Judgment on the basis of *general impression* is one type of interaction in which summation cannot be assumed. The situation is restructured and the separate facts of the case, instead of being separately evaluated, are integrated behind one dominant fact or general impression. In other words, the case is pre-judged on the basis of one dominant factor, and the other factors contribute only by reinforcing, or at least not opposing, the dominant factor. Figure 2 is an attempt to represent this interaction schematically.

The best-known example of this phenomenon is the "halo effect" as it appears in judgments of personality. The halo effect has been widely discussed; it is sufficient here to quote Gordon Allport's (2) comments.

> The halo effect appears with monotonous uniformity in nearly all studies of ratings, and its magnitude is often surprising great. The judge seems intent on reporting his final opinion of the strength, weakness, merit, or demerit of the personality as a whole, rather than on giving as discriminating a rating as possible for each separate characteristic. Whenever the variables have moral connotation the halo effect is larger, for it is a striking fact that a general attitude of approval or disapproval toward the subject colors every single judgment concerning his single vices and virtues. The halo effect is also large when any single variable is not easily observed in action or when it is ill-defined; in such cases the judge substitutes his general impression for the variable that he cannot rate directly.
>
> The halo has considerable theoretical significance. Its existence is proof positive that in perceiving and reflecting upon a personality we rapidly structure our impressions into a self-consistent totality. The structuring is far more rigid and coherent than it should be. Though it dulls our discriminative capacity it demonstrates for us one of the essential characteristics of intuitive knowledge, namely, its tendency toward totalized structures (p. 447).

This concept has worked its way into the technical psychological literature and even into the elementary textbooks, but anyone who looks for clear-cut evidence on the halo effect is certain to be disappointed. The evidence usually given comes from correlations between ratings of personality traits since confusion between the separate traits and the general impression will increase the correlation between the separate traits. Recently the correlations have been subjected to factor analysis. In one instance (30) factor analysis of an industrial merit rating scale of twelve separate characteristics showed that almost all of the variance of the whole scale could be accounted for (after rotation) by one factor, which the authors therefore called "ability to do the present job." The difficulty with such evidence is that correlations between

ratings are ambiguous. It is not easy to decide whether the correlation is the result of a confusion in the mind of the rater or of an objective relationship within the variables being rated. Thorndike (102), in a paper published in 1920, attempted to show that the objective relationship accounts for only part of the correlation, as follows:

> It is known from abundant evidence that technical ability as a flyer is a rather highly specialized quality. Considering the restricted range of the aviation cadets, the correlation between general ability for officer work and technical ability as a flyer could hardly be above .40, without any attenuation. As attenuated by the imperfections of the rater's knowledge of both, it could hardly be above .25. Yet the correlations for the eight raters studied in this respect are .74, .85, .52, .91, .63, .72, .47 and .53, an average of .67. Obviously a halo of general merit is extended to influence the rating for the special ability, or vice versa (p. 27).

This sort of argument is at least plausible and can be taken as proof of the halo effect if one is willing to accept Thorndike's estimate of the objective correlation.

Symonds (99) approached this problem with the partial correlation technique in an attempt to eliminate the effect of the general impression and thus to ascertain its importance. He had two teachers rate their pupils on seven personality traits. The correlations between their ratings ranged from .19 for orderliness to .47 for honesty, with a mean of .39. A composite rating was obtained by adding the seven separate ratings for each pupil, and this composite was taken as each teacher's general impression. He then computed partial correlations of the second order to find the relation between the two ratings of each separate trait with the influence of each teacher's general impression partialled out. The resulting coefficients ranged from $-.04$ to .55, with a mean of .15. The mean difference of .24 is taken as an indication of the halo effect, and Symonds says that "the halo effect which heretofore has been merely assumed is here demonstrated to be a reality." He goes on to compare one trait with another in respect to this measure and tentatively assigns certain reasons why one trait shows more halo than another. These reasons have often been cited.

Yet this argument is open to the same criticism as the others, namely, that the correlation may be in the people rather than in the rating process. If Symonds had begun with objective test scores, a similar set of correlations might have been obtained. The correlations between the separate test scores would be considered reliability coefficients and the correlations between each test and the battery would be considered a measure of the influence of the general factor in each test. The reliability of each test would, of course, be lowered when the general factor is partialled out. Hence it is not necessary to assume any halo effect to account for these results.

Thus it appears that the evidence for the halo effect is rather thin. Although everyone who has had any experience with ratings of personality acknowledges the importance of this phenomenon, it would be well to have objective data on its incidence and amount. The best technique, the writer would suggest, is manipulation of the procedure of judging

rather than manipulation of the data obtained by one procedure. Symonds' (99) suggestion that all persons be rated for one trait at a time ought to reduce the effect of the general impression, and a comparison of the correlations obtained in this way with those obtained by the customary procedure would be quite clear-cut. Such a comparison apparently has not been published.

Other illustrations of, and more satisfactory evidence for judgment on the basis of general impression, come from studies of the "atmosphere effect" in judging syllogisms and from research on attitudes.

Sells, (95) technique in verifying the effect of the atmosphere or general impression was, in contrast to the above, not to try to remove it or weight it, but to use it in predicting judgments. He assumed that judgment of the validity of the conclusion of a syllogism depends on the atmosphere created by the premises as well as the logical relations of the premises. Negative premises set up a negative atmosphere. Affirmative premises set up an affirmative atmosphere. And conclusions in agreement with the atmosphere thus set up are likely to be accepted. There are complications, but specific predictions can be made from the atmosphere of the premises. Sells presented syllogisms with a large percentage of invalid conclusions to a group of educated adults and found that of these invalid conclusions those in agreement with his predictions were accepted much more often than the others.

Evidence of another kind for the importance of the general impression comes from a study by Asch, Block and Hertzman (4) in which the judging procedure was experimentally manipulated. In one part of this study they had college students rank ten professions according to six characteristics, such as intelligence and social usefulness. The correlations among these rankings were positive and of medium size. Next, they introduced a group or majority standard similar to the subjects' own standards by printing on the blanks the mean rankings for intelligence obtained from a similar sample of students. This operated to raise the intercorrelations slightly, presumably by strengthening the subjects' general impression of each occupation. A fictitious standard was introduced in the next experiment, one which was very different from the mean rankings of the previous sample. The intercorrelations produced under the influence of this unconvincing standard were definitely lower and some were negative, presumably because the general impression had been destroyed. Further studies with other kinds of material, such as photographs, names of political figures, and slogans yielded rather clear-cut evidence for the importance of a general attitude underlying judgment of the specific items.

The effect of a stereotype may be considered an example of judgment on the basis of a general impression, certainly so if we accept Edwards' (29) definition: "A stereotype is a stimulus which arouses standardized preconceptions which are influential in determining one's response to the stimulus." This definition is stated in terms of the judging process,

and Edwards recommends its adoption in place of Lippmann's original "standardized picture in the head." (One should remember, however, that what Lippmann was trying to emphasize by his picturesque term was not the judgment process but the astonishing durability of the general preconception in spite of inconsistent specific data.)

A third mode of interaction—which also emphasizes the structure of the situation—can be identified, in which some of the contributing factors are integrated into a *context* or background, and the remaining factors make their contributions within this context. Consider a case at law as an illustration. A man starts a fight on the street. A policeman accuses him of disturbing the peace and brings him before a judge. The judge's decision will take into account the background of the act, the time of day, the events leading up to the act, the occupation of the accused, his reputation, and so on. These facts of the case are not summated, obviously. Certain facts, meanings, implications, connotations, associations, etc. are admitted to judgment. Others are thrown out. The data admitted are organized into a context, and within this context or frame of reference* the act is judged. Other examples need not be cited. The strategy of painting a background against which a proposition or client appears in a favorable light is widely used by public relations counselors, debaters, defense attorneys and political campaign managers. Cantril (17), in constructing a systematic framework for his analysis of social movements, devotes a chapter to "the individual's mental context" wherein he discusses such problems as how the values of the culture are "interiorized" and how the ego gets involved in the context. Figure 3 is intended to illustrate this mode of interaction.

Clear-cut data on the context of judgment are hard to find. Context is usually brought in after the fact as an interpretation of a shift of judgment or of individual differences. But an interesting example of one way in which the context may be changed comes from an attitude study by Asch (3). As one phase of this study he had two samples of students from the same college population rank ten occupations in respect to intelligence, social usefulness, and the like. They were given, "as an illustration," the ranking of politics by a group of 500 college students. For one group this fictitious ranking of politics was 1, for the other it was 10, and the resulting rankings showed the usual suspectibility to majority opinion. But the important point is that, when they were asked what politicians they had in mind while making the judgments, the first group used phrases like "more statesmanship than politics," while the second group spoke of the "usual neighborhood politicians." One can explain

* The term, "context" is used here rather than "frame of reference" as it is narrower in meaning and easier to define. "Frame of reference" has a number of meanings in psychology; in fact it is sometimes used to include phenomena which are discussed in the next section under the heading of "response scale."

such evidence away, to be sure, as *ex post facto* rationalizations for the influence of suggestion. But the statements of the subjects sound too specific for this simple interpretation. They imply that the subjects used the two standards to establish two different contexts within which the occupation of politics is judged. Or, as Asch put it, they "so interpreted the object of judgment as to make the standard reasonable." From a similar experiment on the changing of attitudes by Lewis (74) we can get further illustrations of the operation of context. In interpreting her

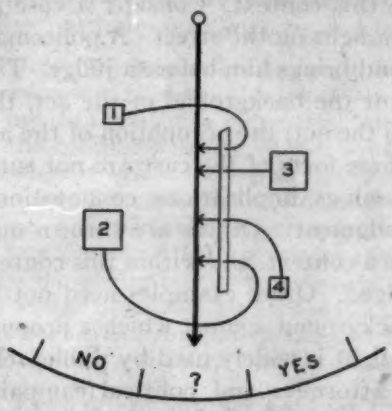

FIG. 3. In the mode of interaction illustrated by this figure the variables of the judgment are assumed to be structured by a context or frame of reference within which they operate.

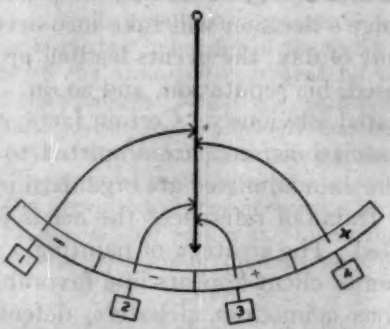

FIG. 4. Illustrating the special case of judgment of simple uniform stimulus objects in which there is no context other than the range of stimuli presented. The context of judgment and the response scale are formally equivalent. An example is the psychophysical judgment by the method of Single Stimuli.

results she says that the source of the statement functions to supply context for the statement, and change of opinion depends on whether "the new pattern thus created seems at least as clear and correct as the old." In Cantril's (17) terminology this illustrates the "individual in pursuit of meaning."

No doubt there are many other ways of structuring and restructuring the facts of the case. Blumenfeld (9, 184–188), in his comprehensive 1931 treatise on judgment prints a tentative schema of insightful operations on the situation to be clarified. His main headings are analysis of the material, combination of independent data, and reconstruction of the perceived situation as a whole. There are several subheadings under each of these. Blumenfeld also emphasizes the role of doubt in changing the structure of the situation, and, in the interests of completeness, one should point out that the deliberate doubtful judge may carry out all kinds of operations on the material at hand in accordance with his train-

ing and motivation. Right here is the line—as the writer interprets the trend of research on the higher mental processes—between judgment and problem solving. When a person who is forced to make a choice from a number of alternative ways of responding, instead of dealing directly with the facts of the case as they are presented to him by his perceptual, affective or memory apparatus, invents some new (to him) organization of the data, devises some new solution, or institutes a search for some other way out of the problematic situation, the process is a more complex one, and goes by the name of problem solving.

II. Dependent Variables of Judgment

The judgment—which concludes the judging episode—may be expressed in action, as when the individual takes the right road rather than the left, or it may be communicated to another person by a word or other symbol, or it may merely be registered in memory as a belief or guide to subsequent action. The experimental evidence deals largely with verbal expressions of judgment, though the relation between words and actions has come in for some discussion in connection with attitude measurement (**85**, pp. 889–912).

It is in accord with the present status of psychological research and also with the notion of judgment as settlement of a problematic situation to note that all judgments are relative, and that they are expressed in relation to alternative categories of response or in relation to response scales. Analysis in these terms is, and has been, very revealing so long as the operations of the psychologist are not confused with the operations of the judge. Judgment in categories is illustrated by a student taking a true-false test or judging the items on an attitude test. The problem solver may judge a hypothesis "plausible." The poet may judge his product "incomplete." (Hence, as noted earlier, the judging process may be a final or an intermediary phase of problem solving.) Adjectives such as heavy, red, true, unpleasant and illegal, when used to express a judgment, are relative in the sense that their use is influenced by the other available categories of response and by the judge's understanding of the limits of the categories. Since judgments are, by definition, made in at least two categories, e.g. X and not −X, and since these categories can be treated as forming a response scale of at least two categories, it is profitable to reverse the history of psychology and consider judgment on a scale as the prototype.

A. The Response Scale

Judgments of stimulus objects one by one in "absolute" units—the method of Single Stimuli—express the position of the objects of thought on a scale. The scale may be a well-known scale, familiar to many, as when judgments are made in inches or minutes, or when handwriting

samples are judged as 5th, 6th or 7th grade quality. Or it may be peculiar to the individual judge, lasting perhaps only for the duration of one experiment, as when he judges weights as #1, #2 and #3 or judges personality traits as −2, −1, 0, +1 and +2. Practical suggestions regarding the use of such rating scales are given in the textbooks on personality and industrial psychology and in general treatments of psychological measurement. Our concern here is the more fundamental question of how these scales are developed and used.

In contrast to context, scales are easy to work with for they are simple and one-dimensional, while the context may be multi-dimensional. Our definition of judgment implies that the response is simpler than the stimulus-situation. But some confusion between the context of judgment and the scale of judgment is possible, since both are often considered in discussions of the frame of reference. This can be clarified by taking up again our previous illustration of the man who was accused of disturbing the peace. Let us say that, within the context which the judge has constructed from the facts of the case, he has judged the man guilty. Now the judge has to pronounce sentence. The severity of the sentence is more than a matter of context. It depends on whether the judge calls the offense *very serious*, *serious*, *minor* or *trivial*, in other words, where the offense falls on the judge's scale of the seriousness of crimes. Thus we see that the context determines what independent variables or factors are considered in the judgment and what force they will have. The scale determines the category of response by which the judgment is expressed. This distinction is clear for complex judgments which are determined by a variety of independent variables. In the special case of judgments of simple uniform material, e.g. a series of weights, there is no context other than the range of the stimuli, hence context and scale amount to the same thing. (See Figure 4.) Such simple material is not suitable for investigating the operation of the context or frame of reference in a complex judgment, but is suitable for analysis of the operation of a response scale.

It is important to note also that a figure-and-ground effect occurs here. The object or act being judged may stand out as figure, and will be attended to and remembered. The objects in the background, the experiences on which the scale of judgment is based are usually taken for granted and are not brought out in the subject's report of the factors by which the judgment was determined.

Experimental investigation of a response scale requires, of course, that the scale be described quantitatively. The quantitative description of a response scale in units of the stimulus variable is easily accomplished by simple calculations based on the frequency of use of the categories along the scale. For example, if the stimuli are lines measured in millimeters and the response categories are *longer* and *shorter*, conventional

psychophysical calculations, as for the method of Constant Stimuli, yield a statistical limen or boundary between these categories in millimeters. If a standard stimulus is used for comparison, this value is called the "point of subjective equality" or PSE. If no standard is used, this point is merely a boundary between adjacent regions of a scale. And, if no standard is used, the method is a general one which can be used to describe a scale of any number of categories in terms of transition points or boundaries between any two adjacent regions. Blumenfeld (**9**, pp. 393–467) has published an exhaustive analysis of scales of two, three, five and ten categories, including computation of limens between categories (Trennfugen), category widths, asymmetry, overlapping and the like. Similarly, any one category, e.g. *equal* or *mellow* or *moderately bitter*, can be separately studied so as to get a quantitative definition of the concept involved in relation to the remaining categories, the instructions and other experimental conditions. Cartwright (**18**) has used such a procedure in determining "ranges of equivalence" for words.

If we raise the question *How does the judge acquire his scale of judgment?* the answer must be that he learns it. Certainly it is not an innate characteristic of the perceptual apparatus. The best data on such learning come, not from the use of a well-established scale of inches or minutes but from experiments in which the subject develops and uses a new scale in a controlled situation. Whether they are asked to rate the pleasantness of color cards on a seven-point scale, or to judge weights as *one*, *two* and *three* (**112**), or to judge the irregularity of a scattering of points on a ten-category scale (**9**), or to judge the meaning of words on an eleven-category scale (**84**), people are able to adapt to such scales quickly. The evidence for this is that they make their judgments readily, using all categories with some degree of consistency and uniformity. The presentation of low stimulus values lowers the scale; high values raise it. Hence, as a first approximation, we may say that, at least with *ad hoc* scales based on experience with a uniform series of stimuli, the response scale is co-extensive with the range of stimuli used in an experiment. Hunt and Volkmann (**55**, p. 88) have put it as follows:

> The position and the width of this scale are determined by the position of the group of stimuli and by the stimulus-range which the group covers. In general, when the group of stimuli is moved up or down upon the stimulus-continuum, the absolute scale moves with it; when the group of stimuli expands or contracts, the scale likewise expands or contracts. There is still variability of judgment, but the general position of the scale is determined. Under these conditions, it seems appropriate to speak of the scale as being *anchored* by the stimuli.

A recent paper by Johnson (**63**) begins at this point and attempts to describe the relation between the experience with the stimulus objects and the scale which is thus learned. He takes the lifting and judging of a weight as a unit of practice with that weight. On the basis of a mathe-

matical discussion of the organization of a scale—starting with what is known about generalization along a stimulus continuum—he assumes that the practice effects may be averaged to determine the location of the center of the scale. Although, in the absence of empirical data, he is forced to assume a questionable equation for the receptor function which delivers the effects which are averaged, his predicted category boundaries for scales of two and of four categories agree quite well with experimental values obtained from a variety of stimulus distributions.

In this investigation all the stimuli were similar. It is another question how similar the stimuli must be in order that their effects will be organized into a common scale of judgment. The question in this form has not been investigated, but the previous discussion of summation may be pertinent. In an incidental way it has been reported (63) that a scale organized around lifting laboratory weights is not displaced noticeably by other kinds of lifting, such as lighting a cigarette and shifting a chair.

The scales of judgment used outside the laboratory show irregularities, e.g. the "round number" effect, here and there. In a psychological scale of warmth the "physiological zero" probably exerts a special influence. There are a number of experiments in which the scale of judgment has been experimentally manipulated so that modifications of the scale and their relation to the presented stimuli can be observed.

Blumenfeld (9, pp. 393–406) was interested in a variety of scale which may have some significance for ratings of personality, namely, a scale which is anchored at one end by an imaginary or ideal limit. His stimuli were scatterings of points to be judged for regularity. There is no apparent limit to the irregularity of the points, but the limit of regularity would be reached when the points all lie in a straight line. He has considerable evidence which emphasizes the importance of this ideal limit. The subjects judged these point figures by imagining a lower limit of zero irregularity and anchored their scales at this end. The category next to the ideal limit was smaller than those farther away. It is likely that our daily value judgments are made on a scale which is anchored at some such ideal reference point. Blumenfeld mentions the concept of normality in this connection. Scales in which the categories are percentages, and other scales which are bounded by zero and unity, are probably anchored at both ends. This end-anchoring would appear to yield an advantage in respect of inter-individual consistency which has not been exploited in practice.

The effect of end-anchoring can be easily demonstrated, and the details can be studied by the use of an extreme magnitude, real or imaginary, as an anchoring stimulus. Presentation of an extreme value extends the scale, hence subsequent judgments are expressed in less extreme categories. This holds for inclinations of lines (91, 109), for pleasantness of colors (55), for weights (91), for the prestige of occupations and the undesirability of certain forms of behavior (80). The effect of the extreme value is a systematic one, which influences all category limens, usually broadening the categories (80, 91). Since this phenomenon has been demonstrated in judgments on many kinds of material, perceptual, affective and abstract, it may be taken as a general principle of judgment (53, 80).

Ratings are often made with standards or models in view, as in grading fruit

and rating handwriting. These standards act as anchoring values, tieing down the scale at certain points along the stimulus continuum. Blumenfeld (9, pp. 436–448) compared the judgments which result when such standards are given as types, each one representing the type or center of a category, and when they are given as limits or boundaries of the categories. His results lead him to favor the use of limits rather than types but, as he points out, the evidence is not clear-cut. In the light of the prevailing custom of designing rating scales with descriptive phrases at the centers of the categories this is still an important practical question.

Not only can the ends of a scale be extended, as we have seen, but the scale can be condensed in some regions and expanded in others by suitable stimulus presentations (75). Variability of judgment depends in part upon stimulus density, i.e. the number of stimuli per unit of stimulus range (54, 110).

It is obvious from the above that there is a definite advantage in treating the categories of judgment as regions of a scale; hence we shall mention only one illustration of how this modern enlightenment clears up an old psychological problem, namely, the problem of the judgment of emotional expression. It has been recognized for some time that the words used in naming facial expressions are not logically exclusive and that measures of accuracy of naming based on that assumption are too rigid. Attempts were made to solve the problem by grouping the names into several broad categories, but this is merely a logical procedure which does not yield a unique solution. Woodworth (115, pp. 249–252) went a step farther and aligned the groups into a scale of six broad categories. The alignment makes sense in respect to the scattering of judgments and permits the calculation of a defensible correlation between stimulus and response. Schlosberg (94) has analyzed judgments obtained by the use of this scale in more detail and finds much overlap between the sixth category and the first. This leads to the conclusion that the scale is a recurrent one, probably elliptical rather than circular. Thus clarity is brought into a chaotic field, and the experimental results can now be given a rational treatment.

The *comparative judgment*, as it is usually carried out, is best understood as a special case of judgment on a scale. The scale is established by the series of comparison stimuli; the standard is an anchoring stimulus which holds down the middle of the scale. Although comparison with the standard is emphasized by the customary instructions, several investigators have called attention from time to time to the "absolute" impression obtained from a single stimulus as a result of its position in the series of comparison stimuli. Hollingworth's (49) work on "the inaccuracy of movement," in which he found that constant errors in the reproduction of extents by hand movements increased with the extension of the series limits, led him to emphasize the "central tendency" of the series in the determination of the indifference point. Ipsen (56), working with a series of Sander's figures, had his subjects estimate the ratio of one diagonal to the other and found that the equivalence value

did not coincide with either central tendency (mean or median) of the objective stimulus values, but as the mean varied the equivalence value varied in the same direction. It was this effect of the series, confused in the comparative judgment with the effect of the standard, which led to the more recent studies by the method of Single Stimuli mentioned above.

Another old psychological problem which has been settled by recent knowledge of the scale of judgment is the problem of the middle category in the comparative judgment: Shall the subject be permitted to say *equal* and *doubtful*, or shall a definite choice one way or the other be required? The early discussions of this topic centered around the distinctness of the impression of equality. When this criterion was set aside and the *equal* response treated as the middle category of a three-category scale, the whole problem was clarified. The problem became one of the variability of the width of the middle category and the vulnerability of this width to attitudinal changes. Dependable quantitative data are available on the question when put in this form (**18**, **35**), and the methodological implications have been adequately discussed (**115**, pp. 421–425).

If we go further and attempt to apply our knowledge of the organization of a scale to the comparative judgment in a systematic way, we come out with some interesting results. Suppose we have five comparison stimuli, weights of 92, 96, 100, 104 and 108 grams, to be compared with a standard of 100 grams. As the experiment proceeds, the subject will soon be judging each comparison stimulus in relation, not to the standard alone, but to the central tendency of the scale built up by experience with all the weights. Now, if we assume that the central tendency of the scale can be computed by averaging the central effects of experience with the stimuli—an assumption which has some empirical support (**63**)—and, furthermore, if we make the usual assumption that the receptor function is a logarithmic one, it will turn out that the PSE will always be lower than the standard stimulus because the standard is set at the arithmetic mean of the series and the geometric mean is always lower than the arithmetic mean. In other words, a negative constant error will occur whenever a scale is organized around an average of the effects delivered by a receptor function of decreasing slope. Negative constant errors are commonly found in psychophysical judgments, most consistently in judgments of weights and of intensities of sound. It is noteworthy that the receptor function for both of these modalities is usually assumed to be a logarithmic one, or something similar.

It is easy to compute the constant error by these principles. Instead of assuming a logarithmic receptor function and computing a geometric mean we shall assume the modified logarithmic function previously found useful for lifted weights (**63**). If the standard is not judged, but merely observed, it can be left out of the calculations. Computing the central tendency, then, for the

five stimuli of our above example we get 98 grams, which will be the PSE. If the standard is included in the calculations once for each comparison stimulus, we compute the central tendency of ten stimuli, which comes out as 99 grams. These two values may be taken as the limits within which the PSE will fluctuate in accordance with the relative attention value of the standard. The PSE usually obtained for weights of this range does vary around these limits. It usually decreases a little, as one would expect if the attention value of the standard decreases in comparison to the series effect. If we assume that, for the experiment as a whole, the standard has an attention value half that of a comparison stimulus, we can weight the calculations accordingly and get a single value for the PSE. In this example it will be 98.6 grams.

Wever and Zener (112) used a series of weights of 88, 92, 96, 100 and 104 grams with the standard at 100 instead of at the midvalue of 96. A general theory should predict the results from this unusual arrangement, however. If we carry through the computations in the above manner, we get a PSE for this arrangement of 96.1. The mean PSE for their six subjects if 96.7. They also publish results for two subjects with a wider range of comparison stimuli and the same unusual standard. Our method of computation gives a PSE of 95.2 for such a series, and the two values given by Wever and Zener are 95.3 and 96.7. If we combine these two values with eight PSE's published by Fernberger (36) for the same series, we get a mean PSE of 95.2 grams. This agrees, as well as one could wish, with the theoretical value.

It turns out, therefore, from this discussion of a constant error in the comparative judgment—which is often called the "time error"—that there is a genuine advantage in treating the comparative judgment systematically as a special case of judgment in reference to a scale. We hasten to add, however, that there is more to the "time error" than this. A negative constant error, which is properly called a "time error," is produced by the order of presentation of standard and variables and is covered up by the customary balanced method pf presentation. It may be brought out by comparing the judgments when the comparison stimulus is given before the standard and when it is given after the standard. Woodrow (114) attacked the whole question of the effects of the standard by systematically using a wide range of different standards. He found that he could separate the obtained errors into two parts, and could account for these errors by postulating two tendencies: "first, one producing a negative error throughout, independently of the weight of the standard; and secondly, one causing increasingly positive or negative errors with increase in the difference between the weight of a standard and the middle or average weight of all the standards" (p. 403).

The first tendency is not clarified by the present discussion. It is presumably a neurological phenomenon, and has been attributed to the "fading" or "sinking" of the effects of stimulation. It is the second tendency, the "central tendency," empirically observed throughout the history of psychophysics and attributed to such principles as "set" and "assimilation," which may now be supplied—if the above discussion is correct—with a rational explanation and a method of computation.

B. Confidence, Time and Difficulty

Looking at judgment as settlement of a problematic situation we can interpret a report of the confidence felt in a judgment as a report of how problematic the situation remains after the judgment is delivered. In fact judgment can be defined in terms of confidence in the result (**50**), or at least differentiated from certain other processes. Learning experiments can be arranged so that confidence increases as the situation becomes less problematic (**64, 103**).

Confidence and time of judgment are easily recorded. The scale for reporting confidence runs from zero confidence, "a pure guess," to complete confidence, "100% certainty." Frequency distributions of such reports are J- or U-shape (**96**, pp. 26–27) because the scale is anchored at the two extremes. Occasionally the scale may be anchored in the middle also (**59**, pp. 39–41). These irregularities in the confidence scale raise difficulties in the quantitative treatment of confidence reports, but quantitative data are available and their consistency argues for their validity.

Many experimenters at various times have asked for reports of confidence, usually after comparative judgments on psychophysical material, and it is now well recognized that confidence in the two-category judgment increases with an increase in the physical difference between the stimuli being compared. It has recently been determined (**37, 59**) that the relation between confidence and the stimulus variable is an ogive, similar in shape to the usual relative frequency function but of more gradual slope. Volkmann has similar curves (unpublished) for the three-category judgment. Johnson (**59**) makes the generalization, suggested by Volkmann, that *confidence increases as a function of distance from a category threshold.* In general, any change in the stimulus material which increases accuracy of judgment increases confidence in the judgment (**41, 43, 59**). This follows from the statement above that the confidence report is an indication of how problematic the situation remains. Constant errors complicate this relation sometimes (**59**, pp. 41–44). Under some conditions the relation is reversible: Those judgments given with most confidence are the most accurate (**15**). Curiously, speeding up the judgments by instructions to emphasize speed rather than accuracy has no significant effect on confidence, at least not for judgments of weights (**41**) or of lengths of lines (**37, 59**).

Time of judgment has been studied occasionally—perhaps because of the analogy with reaction time, although in judgment the emphasis is on accuracy rather than speed and the times are much longer. Working with brightness comparisons Kellogg (**66**) plotted time functions on the stimulus variable separately for each category of judgment and thus showed that "the psychometric time curves for all categories tend to be inversions of the psychometric frequency curves." He also plotted time

on the stimulus variable for all categories of response. This curve, of course, is highest near the point of objective stimulus equality and decreases on both sides of this point, as stimulus-difference increases. On the basis of these curves Kellogg could make the generalization that "within any given category of judgment, the more the stimuli deviate from the ideally perfect condition typified by the category name, the longer the judgment-times tend to become." Similar curves for other kinds of material and similar generalizations have since been published (**18, 59**). Since it is difficult to locate the "ideally perfect condition" on the scale, it is preferable to state this generalization in relation to a point which can be located, as follows: *Judgment time decreases as a function of distance from a category threshold* (**59**). This relation is a general one which holds for a scale of affective value as well as physical value (**6, 21, 25**).

Cartwright (**18, 19**) has attempted to interpret results of this sort in terms of "differentiation of the phenomenal field"—a Lewinian concept. Instead of "category threshold" he speaks of a "border of a range of equivalence," and he presents data from a variety of ingenious experiments to demonstrate how a range of equivalence can be broadened and narrowed by the experimental conditions. Considering a range of equivalence as a region of the phenomenal field Cartwright goes on to say that the subject, in his attempt to be right, will set up certain forces related to the alternatives in the phenomenal field, and that decision time is a resultant of the conflict between these forces. Since conflict arises and judgment is prolonged when the stimulus falls upon a boundary of a region in the phenomenal field, judgment time will depend upon the differentiation of the phenomenal field.

In the writer's opinion Cartwright falls into a minor error when he confuses the number of categories of the scale with the differentiation of the field. He cites Kellogg's experiment—the one mentioned above—in which the mean judgment time for the three-category judgment was 10% greater than for the two-category judgment as having general implications. Suppose, however, that the subject is judging a series of three stimuli, evenly spaced, instead of the seven that Kellogg used. It is quite likely that the two-category judgment would be the slower one on the average with such a stimulus range—and with many other ranges which could be set up. Differentiation of the field must be defined in terms of what the subject is trying to do.

In a later series of papers Cartwright and Festinger (**20, 37, 38**) put forth a mathematical theory of these conflicting forces which determine the decision time. The forces are not quantitatively specified except in the sense that the resultant of two opposing forces is normally distributed in time around a mean value. To prevent a decision from occurring on the basis of only a slight imbalance of the opposing forces the authors introduce "restraining forces" with normal distribution about a mean value. A decision occurs only when the resultant of the opposing forces is greater than the restraining forces of the moment. The probability of different categories of decision can now be written in terms of these postulated forces. Taking several convenient values for the mean restraining force they draw theoretical curves for relative frequency of response and for decision time as functions of mean resultant force. They are thus able to derive several semi-quantitative tendencies of judgment which check quite well with a variety of empirical data. A more precise check necessitates tieing

the hypothetical resultant force to empirical data in some way. This is done by the statement—called an operational statement—that confidence is a linear function of mean resultant force and may be used as a measure of it. Hence they come to a plot of the relation between time of judgment and confidence for different magnitudes of the restraining force (imposed by different instructions).

It is difficult to evaluate the fit of the theoretical curves which Cartwright and Festinger draw over their empirical data. Certainly in Figures 1a, 1b, 1c, 2b, 3 and 4 (**38**) the departures from the theoretical curves are not random. They suggest a peak near the point of zero confidence, i.e. near the category threshold, similar to curves previously published (**59**, p. 42), rather than the flat-top curve which the theory requires. But this is a detail; on the positive side is the important fact that they can calculate constants from the results obtained under one kind of instructions and find that they hold for other kinds of instructions. All in all, while they have not given their theory a rigid test—as they claim—they have taken a big step in drawing together a number of experimental findings into a coherent theory, showing that a quantitative theory about the higher mental processes is possible and profitable.

The relation between the time required for judgment and the confidence expressed in the judgment has been taken seriously, and we can make the generalization for any person judging a wide range of any kind of material, i.e. both large and small stimulus-differences, that those judgments which are given most quickly will be given with most confidence. Seward (**96**) studied this relation by the correlation technique, using recognition judgments of fancy papers. Her correlations for many subjects ranged from .37 to .81 with a median of .65. This technique, of course, assumes a linear relationship. Two attempts have been made to determine just what the form of the relation is. Volkmann (**108**), on the basis of judgments of inclinations of lines, published a tentative equation in hyperbolic form. Johnson's (**59**) equation, which fits the data on judgments of three kinds of material fairly well, is in logarithmic form: "As doubt or uncertainty increases arithmetically, judgment time increases geometrically." The departures from these simple equations are chiefly in the region of the threshold when the large degree of doubt may lead to a complication in method of reaching a decision, e.g. "a giving-up." It is still a question whether doubt is the cause of long judgment times, or the result.

Although we know in a general way that confidence and time are each related to the difficulty of a judgment, precise analysis is lacking because of the difficulty of getting accurate measures of difficulty. Any careful analysis requires measures of difficulty which are valid for the person making the judgments. In the absence of this measure difficulty has been measured in terms of a social criterion, namely, the percentage of persons failing an item. And it is known that under some conditions felt difficulty is related to this social criterion in a logarithmic way (**45**). Items which are more difficult in this sense are judged with lower confi-

dence (**47**), with more muscular activity (**26**), and with a greater drop in palmar skin resistance (**39**).

It is a safe generalization that several dependent variables are related to distance along the scale from the category threshold: confidence, time, difficulty, and effort as measured by muscle activity and skin resistance. Therefore, when one category of judgment is given with more confidence (**1, 60, 96, 98**) or less time (**21, 66, 73, 86, 96**) or greater GSR (**21, 73**) than another, an adequate interpretation of the results is not possible unless this distance from the category threshold is considered. The old problem of the relation between atypical opinion and confidence has been taken up again from this point of view (**60**).

Blumenfeld (**9**, pp. 189–192, 484–488) has been particularly interested in the vacillation or oscillation (Schwanken) which enters the judging process along with doubt. His interpretation is based on comments of his subjects while judging his point figures and his observations of their behavior. To illustrate this vacillation which occurs when neither opposing force is strong relative to the other he refers frequently to the analogy with a pendulum. Since some sort of vacillation is commonly mentioned in the literature on judgment, the pendulum analogy is included in the figures which accompany this article.

According to Blumenfeld we incline now to one side and now to the other, under the influence of opposing forces of approximately the same strength, and finally one attains a preponderance. The pendulum system is damped somehow, and Blumenfeld suggests the time of judgment as a measure of the damping of the energy system under certain simple conditions. Actually, while the swaying continues, the forces are changing as a result of incoming perception, accidents or information. Reconstructions are occasionally so made that the whole energy system is deformed. If doubt continues, and the activity becomes annoying, the judgment may be evaded, a frivolous decision may be made, or the categories of response may be changed so as to make judgment easier.

Doubt may be objectively observed, according to Blumenfeld. The lips are shoved forward and backward. The hands beat time. The fingers make drumming movements. The head or trunk bends this way and that. The brow is wrinkled. The lips bite together. He measured the time of oscillation for a few subjects as they expressed it by tapping. The times varied from three to seven seconds. He suggests the possibility of a personal constant of decision analogous to the law of small pendulum oscillations. Changes, after the decision has once been announced, are made when doubt is great. In fact Blumenfeld has some examples of rapid forgetting of the announced judgment, presumably because it was immediately reversed. And some of his subjects said that the judgment had been "extorted" from them, probably because they happened to make the judgment just at the moment when the pendulum was at the limit of its swing.

On this topic Blumenfeld's objective evidence is quite casual. He has relied mainly on introspective data, which are as suspect here as elsewhere, especially in respect to a report of time relations. No one would deny that such vacillation occurs, and there is some objective evidence that judgment times are longer for those choices during which the sub-

the hypothetical resultant force to empirical data in some way. This is done by the statement—called an operational statement—that confidence is a linear function of mean resultant force and may be used as a measure of it. Hence they come to a plot of the relation between time of judgment and confidence for different magnitudes of the restraining force (imposed by different instructions).

It is difficult to evaluate the fit of the theoretical curves which Cartwright and Festinger draw over their empirical data. Certainly in Figures 1a, 1b, 1c, 2b, 3 and 4 (**38**) the departures from the theoretical curves are not random. They suggest a peak near the point of zero confidence, i.e. near the category threshold, similar to curves previously published (**59**, p. 42), rather than the flat-top curve which the theory requires. But this is a detail; on the positive side is the important fact that they can calculate constants from the results obtained under one kind of instructions and find that they hold for other kinds of instructions. All in all, while they have not given their theory a rigid test—as they claim—they have taken a big step in drawing together a number of experimental findings into a coherent theory, showing that a quantitative theory about the higher mental processes is possible and profitable.

The relation between the time required for judgment and the confidence expressed in the judgment has been taken seriously, and we can make the generalization for any person judging a wide range of any kind of material, i.e. both large and small stimulus-differences, that those judgments which are given most quickly will be given with most confidence. Seward (**96**) studied this relation by the correlation technique, using recognition judgments of fancy papers. Her correlations for many subjects ranged from .37 to .81 with a median of .65. This technique, of course, assumes a linear relationship. Two attempts have been made to determine just what the form of the relation is. Volkmann (**108**), on the basis of judgments of inclinations of lines, published a tentative equation in hyperbolic form. Johnson's (**59**) equation, which fits the data on judgments of three kinds of material fairly well, is in logarithmic form: "As doubt or uncertainty increases arithmetically, judgment time increases geometrically." The departures from these simple equations are chiefly in the region of the threshold when the large degree of doubt may lead to a complication in method of reaching a decision, e.g. "a giving-up." It is still a question whether doubt is the cause of long judgment times, or the result.

Although we know in a general way that confidence and time are each related to the difficulty of a judgment, precise analysis is lacking because of the difficulty of getting accurate measures of difficulty. Any careful analysis requires measures of difficulty which are valid for the person making the judgments. In the absence of this measure difficulty has been measured in terms of a social criterion, namely, the percentage of persons failing an item. And it is known that under some conditions felt difficulty is related to this social criterion in a logarithmic way (**45**). Items which are more difficult in this sense are judged with lower confi-

dence (47), with more muscular activity (26), and with a greater drop in palmar skin resistance (39).

It is a safe generalization that several dependent variables are related to distance along the scale from the category threshold: confidence, time, difficulty, and effort as measured by muscle activity and skin resistance. Therefore, when one category of judgment is given with more confidence (1, 60, 96, 98) or less time (21, 66, 73, 86, 96) or greater GSR (21, 73) than another, an adequate interpretation of the results is not possible unless this distance from the category threshold is considered. The old problem of the relation between atypical opinion and confidence has been taken up again from this point of view (60).

Blumenfeld (9, pp. 189–192, 484–488) has been particularly interested in the vacillation or oscillation (Schwanken) which enters the judging process along with doubt. His interpretation is based on comments of his subjects while judging his point figures and his observations of their behavior. To illustrate this vacillation which occurs when neither opposing force is strong relative to the other he refers frequently to the analogy with a pendulum. Since some sort of vacillation is commonly mentioned in the literature on judgment, the pendulum analogy is included in the figures which accompany this article.

According to Blumenfeld we incline now to one side and now to the other, under the influence of opposing forces of approximately the same strength, and finally one attains a preponderance. The pendulum system is damped somehow, and Blumenfeld suggests the time of judgment as a measure of the damping of the energy system under certain simple conditions. Actually, while the swaying continues, the forces are changing as a result of incoming perception, accidents or information. Reconstructions are occasionally so made that the whole energy system is deformed. If doubt continues, and the activity becomes annoying, the judgment may be evaded, a frivolous decision may be made, or the categories of response may be changed so as to make judgment easier.

Doubt may be objectively observed, according to Blumenfeld. The lips are shoved forward and backward. The hands beat time. The fingers make drumming movements. The head or trunk bends this way and that. The brow is wrinkled. The lips bite together. He measured the time of oscillation for a few subjects as they expressed it by tapping. The times varied from three to seven seconds. He suggests the possibility of a personal constant of decision analogous to the law of small pendulum oscillations. Changes, after the decision has once been announced, are made when doubt is great. In fact Blumenfeld has some examples of rapid forgetting of the announced judgment, presumably because it was immediately reversed. And some of his subjects said that the judgment had been "extorted" from them, probably because they happened to make the judgment just at the moment when the pendulum was at the limit of its swing.

On this topic Blumenfeld's objective evidence is quite casual. He has relied mainly on introspective data, which are as suspect here as elsewhere, especially in respect to a report of time relations. No one would deny that such vacillation occurs, and there is some objective evidence that judgment times are longer for those choices during which the sub-

ject looked back and forth between the alternatives (68), but this phenomenon is more parsimoniously interpreted by statistical principles, i.e. in terms of the cumulation of variations in the effects of the stimulus variables, or of a "random distribution of distracting stimuli" (70), or of variations in the opposing forces and the restraining forces (20).

III. Individual Differences in Judgment

Studies of individual differences have not often been directed toward judgment as distinct from other intellectual processes. Individual differences in attitudes, for instance, have usually been related to cultural background, personality and the like, only occasionally and incidentally to the details of judging. There may be significant individual differences, however, in any of the aspects of judgment discussed in this paper. What are called "types of thinking" may upon investigation turn out in many cases to be merely differences in the weight attached to various kinds of material. Presumably the theoretically-minded man—if there is such a man—weights a general principle heavier in making up his mind than the practically-minded man does. In the relative weight carried by emotional and rational factors one would expect to find large differences between cultures and from time to time within the intellectual history of one culture. For example, a careful reader of Curti's recent *Growth of American Thought* (24) could plot the long-time trends in this weighting over a period of three centuries. The more skillful advertisers and propagandists recognize such differences and adjust the rationality of their appeals to the level of their audiences. Individual differences in the intrusion of affective material into an abstract judgment are tested on the Watson-Glaser Test of Critical Thinking (42).

Suggestibility has been studied for some time, and adequate summaries of the literature are available (8, 85). One of the chief topics of investigation has been the question of generality versus specificity. In the present paper we have seen that there are many principles of judgment which operate independently of the material on which the judgments are made. This would lead one to expect some degree of consistency of suggestibility from one kind of judgment to another. On the other hand it has been necessary to treat suggestion in two ways, in terms of the weight attached to a source, and in terms of a restructuring of a situation. Putting these two hints together one would expect to find a moderate degree of generality on tests of suggestibility, and this is in fact what the latest summaries indicate. Subsequent papers support this conclusion (23, 34, 90).

Consistent individual variations in the organization and use of a scale have occasionally been reported. Mosier (84) noted in an incidental way in his "psychometric study of meaning" that there was among his subjects "a tendency to mark words either at the extremes, or

toward the neutral value." Guilford and Jorgensen (46), when analyzing frequency distributions of affective judgments on a nine-category scale, observed "that for the same individual the shape of the curve is decidedly constant whether he is rating colors or color combinations." Examining the use of a seven-point scale in an attitude study Osgood (86) discovered that some people used 1 and 7 almost exclusively, others used only 1, 4, and 7, while others used the entire scale. "Preliminary analysis of these data indicates that these differences are related to occupation, education, and intelligence, the more critical thinkers making a more discriminatory use of the entire scale." If a tendency of this kind can be established with certainty, it would be a matter of decisive importance in interpreting the results of experiments using ratings. For instance Singer and Young (97) found correlations as high as .53 between affective judgments on various kinds of material, a correlation which some would take as evidence for a general hedonic factor, but they preferred to explain it as due to "some consistent manner of using the rating scale." At present both interpretations are plausible. Certainly differences in generalizing ability enter into such judgments, but how they would show up in the judgments is not known.

Individual differences in confidence in a judgment are easily and reliably measured (59, 60, 61, 67), and some evidence from intercorrelations on generality of confidence in judgments of different kinds of material has been reported (58, 59, 60, 61, 67, 106). The amount of generality under favorable conditions is indicated by mean intercorrelations in the .50's and .60's. Bi-factor analysis (61) of confidence scores on eight vocabulary tests allotted 48% of the variance to the general factor. The general factor of vocabulary achievement accounted for only 27% of the variance.

Klein and Schoenfeld (67) repeated Johnson's (59) experiment on generality of confidence *with* and *without* emphasis on the personal significance of the tasks to the subjects. In the first condition, which they call "ego-involvement," the mean intercorrelation was .45; in the second it was .22. This is an imporatnt result. It ties up with the recent attempts to increase the validity of personality tests by the use of the stress situation and with a recent discussion of studies of level of aspiration by Irwin (57). Irwin noted that levels of aspiration, or expectation, were more closely related to actual performance for "realistic" than for "unrealistic" situations, and that the figures for generality of aspiration are higher for the unrealistic than for the realistic situations. If we examine the nature of the situations and the instructions given, we can equate realistic with objective or impersonal and unrealistic with personal and arrive at a generalization which applies to the discussions of both Irwin and Klein and Schoenfeld: When the tasks are of personal significance, with a high degree of ego-involvement, both expectation

before the performance and confidence after the performance will show a higher degree of generality from one task to another than when the tasks are impersonal. In impersonal tasks both expectation and confidence will be more closely related to the objective situation. The reason for this is that reports of both expectation and confidence are judgments, and the objective situation differs from one task to another while the ego factors which contribute to a judgment of performance level remain relatively constant.

In the mental-test situation individual differences in confidence in a judgment are not closely related to intelligence or performance (**43**, **61**, **106**). Confidence enters the picture in another way, however. Gritten (**44**) has shown that, when students taking a multiple-choice test are told to answer only the questions they are sure of, i.e. not to guess, the more confident ones will attempt more items and will get more right. The conventional correction—which might be called a correction for variations in confidence—is fairly adequate in removing the advantage.

On attitude tests an extremist is likely to express his opinions with relatively high confidence (**1**, **16**, **31**, **60**), partly because his position on most issues is far from his category threshold. He therefore has less need to deliberate than the troubled middle-of-the-roader. A group is more confident (**60**), as a group, and also faster (**86**), in dealing with those issues on which there is a majority, and will deliberate longer before taking a position on neutral statements (**21**). Similarly, group agreement in a judgment correlates with reported ease of the individuals making the judgment (**65**).

The generality of speed of judgment has attracted some attention. Since judgment is slow, impeding the "congenial pace" of the individual, one would expect that some people would be more eager, as a general rule, to bring judgment to a close than others. Blumenfeld's suggestion of a personal constant of vacillation has been mentioned. Early empirical studies (**12**, **107**) were largely negative for generality. Symonds (**100**) has summarized the research on the speed-of-decision items on the Downey Will-Temperament Test. Average intercorrelations ranged from .25 to .60. Johnson (**59**) used more adequate timing and found rather high intercorrelations: .79, .84 and .91. He used only three tasks, however. The conclusion for speed is the same as for confidence: Present results indicate at least a moderate degree of consistency from one kind of judgment to another.

BIBLIOGRAPHY

1. ALLPORT, F. H., & HARTMAN, D. A. The measurement and motivation of atypical opinion in a certain group. *Amer. Pol. Sci. Rev.*, 1925, **19**, 753–763.
2. ALLPORT, G. W. *Personality: A Psychological Interpretation.* New York: Henry Holt, 1937.
3. ASCH, S. E. Studies in the principles of judgments and attitudes: II. Determination

of judgments by group and ego standards. *J. soc. Psychol.*, 1940, **12**, 433–465.
4. Asch, S. E., Block, H., & Hertzman, M. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment. *J. Psychol.*, 1938, **5**, 219–251.
5. Baker, L. E. The influence of subliminal stimuli upon verbal behavior. *J. exp. Psychol.*, 1937, **20**, 84–100.
6. Barker, R. G. Time required for 'real' and hypothetical choices. *Psychol. Bull.*, 1939, **36**, 627.
7. Beebe-Center, J. G. *Pleasantness and Unpleasantness.* New York: Van Nostrand, 1932.
8. Bird, C. *Social Psychology.* New York: Appleton-Century, 1940.
9. Blumenfeld, W. Urteil und Beurteilung. *Arch. ges. Psychol.*, 1931, dritter Ergänzungsband.
10. Boring, E. G. A chart of the psychometric function, *Amer. J. Psychol.*, 1917, **28**, 465–470.
11. Bressler, J. Illusion in the case of subliminal stimulation. *J. gen. Psychol.*, 1931, **5**, 244–251.
12. Bridges, J. W. An experimental study of decision types and their mental correlates. *Psychol. Rev. Monogr.*, 1914, **17**.
13. Brunswik, E. Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, **56**.
14. Burtt, H. E. & Falkenburg, D. R., Jr. The influence of majority and expert opinion on religious attitudes. *J. soc. Psychol.*, 1941, **14**, 269–278.
15. Cady, V. M. The estimation of juvenile incorrigibility. *J. Delinq. Monogr.*, 1923, No. 2.
16. Cantril, H. The prediction of social events. *J. abnorm. soc. Psychol.*, 1938, **33**, 364–389.
17. Cantril, H. *The Psychology of Social Movements.* New York: Wiley, 1941.
18. Cartwright, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196.
19. Cartwright, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.
20. Cartwright, D., & Festinger, L. A quantitative theory of decision. *Psychol. Rev.*, 1943, **50**, 595–621.
21. Chant, S. N. F., & Salter, M. D. The measuring of attitudes toward war and the galvanic skin reflex. *J. educ. Psychol.*, 1937, **28**, 281–289.
22. Coffin, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monogr.*, 1941, No. 241.
23. Cronbach, L. J. Studies of acquiescence as a factor in the true-false test. *J. educ. Psychol.*, 1942, **33**, 401–415.
24. Curti, M. E. *The Growth of American Thought.* New York: Harper, 1943.
25. Dashiell, J. F. Affective value-distances as a determinant of esthetic judgment-times. *Amer. J. Psychol.*, 1937, **50**, 57–67.
26. Davis, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, **23**, 141–158.
27. Dewey, J. *Logic: A Theory of Inquiry.* New York: Henry Holt, 1938.
28. Eastman, M. *Enjoyment of Laughter.* New York: Simon and Schuster, 1936.
29. Edwards, A. L. Studies of stereotypes: I. The directionality and uniformity of responses to stereotypes. *J. soc. Psychol.*, 1940, **12**, 357–366.
30. Ewart, E., Seashore, S. E., & Tiffin, J. A factor analysis of an industrial merit rating scale. *J. appl. Psychol.*, 1941, **25**, 481–486.
31. Ewing, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.

32. FERGUSON, L. W. The influence of individual attitudes on construction of an attitude scale. *J. soc. Psychol.*, 1935, **6**, 115–117.
33. FERGUSON, L. W. Comparison of scale values from the method of equal appearing intervals and paired comparisons method. *J. gen. Psychol.*, 1940, **23**, 431–435.
34. FERGUSON, L. W. An analysis of the generality of suggestibility to group opinion. *Character & Pers.*, 1944, **12**, 237—243.
35. FERNBERGER, S. W. The use of equality judgments in psychophysical procedures. *Psychol. Rev.*, 1930, **37**, 107–112.
36. FERNBERGER, S. W. On absolute and relative judgments in lifted weight experiments. *Amer. J. Psychol.*, 1931, **43**, 560–578.
37. FESTINGER, L. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *J. exp. Psychol.*, 1943, **32**, 291–306.
38. FESTINGER, L. Studies. II. An empirical test of a quantitative theory of decision. *J. exp. Psychol.*, 1943, **32**, 411–423.
39. FREEMAN, G. L., & GIESE, W. J. The relationship between task difficulty and palmar skin resistance. *J. gen. Psychol.*, 1940, **23**, 217–220.
40. FREEMAN, G. L., & SHARP, L. H. Muscular action potentials and the time-error function in lifted weight judgments. *J. exp. Psychol.*, 1941, **29**, 23–36.
41. GARRETT, H. E. A study of the relation of accuracy to speed. *Arch. Psychol.*, 1922. No. 56.
42. GLASER, E. M. An experiment in the development of critical thinking. *Teach. Coll. Contr. Educ.*, 1941, No. 843.
43. GREENE, E. B. Achievement and confidence on true-false tests of college students. *J. abnorm. soc. Psychol.*, 1929, **23**, 467–478.
44. GRITTEN, F., & JOHNSON, D. M. Individual differences in judging multiple-choice questions. *J. educ. Psychol.*, 1941, **32**, 423–430.
45. GUILFORD, J. P., & COTZIN, M. Judgment of difficulty of simple tasks. *Amer. J. Psychol.*, 1941, **54**, 38–52.
46. GUILFORD, J. P., & JORGENSEN, A. P. Some constant errors in ratings. *J. exp. Psychol.*, 1938, **22**, 43–57.
47. HERTZMAN, M. Confidence ratings as an index of difficulty. *J. exp. Psychol.*, 1937, **21**, 113–119.
48. HINCKLEY, E. D. The influence of individual opinion on construction of an attitude scale. *J. soc. Psychol.*, 1932, **3**, 283–296.
49. HOLLINGWORTH, H. L. The inaccuracy of movement. *Arch. Psychol.*, 1909, No. 13.
50. HOLLINGWORTH, H. L. The definition of judgment. *Psychol. Rev.*, 1925, **32**, 337–361
51. HOUSEHOLDER, A. S. A neural mechanism for discrimination. *Psychometrika*, 1939, **4**, 45–58.
52. HOUSEHOLDER, A. S. A neural mechanism for discrimination: II. Discrimination of weights. *Bull. math. Biophysics*, 1940, **2**, 1–13.
53. HUNT, W. A. Anchoring effects in judgment. *Amer. J. Psychol.*, 1941, **54**, 395–403.
54. HUNT, W. A., & FLANNERY, J. Variability in the affective judgment. *Amer. J. Psychol.*, 1938, **51**, 507–513.
55. HUNT, W. A., & VOLKMANN, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.
56. IPSEN, G. Über Gestaltaufassung. Erörterung des Sanderschen Parallelogramms. *Neue psychol. Stud.*, 1926, **1**, 167–278.
57. IRWIN, F. The realism of expectations. *Psychol. Rev.*, 1944, **51**, 120–126.
58. JERSILD, A. Determinants of confidence. *Amer. J. Psychol.*, 1929, **41**, 640–642.
59. JOHNSON, D. M. Confidence and speed in the two-category judgment. *Arch. Psychol.*, 1939, No. 241.

60. JOHNSON, D. M. Confidence and the expression of opinion. *J. soc. Psychol.*, 1940, **12**, 213–220.
61. JOHNSON, D. M. Confidence and achievement in eight branches of knowledge. *J. educ. Psychol.*, 1941, **32**, 23–36.
62. JOHNSON, D. M. A modern account of problem solving. *Psychol. Bull.*, 1944, **41**, 201–229.
63. JOHNSON, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34**, 425–436.
64. JOHNSON, D. M., & RHOADES, C. Measurement of a subjective aspect of learning. *J. exp. Psychol.*, 1941, **28**, 90–92.
65. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The role of form in visual responses to music. *J. gen. Psychol.*, 1942, **26**, 199–222.
66. KELLOGG, W. N. Time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, **43**, 65–86.
67. KLEIN, G. S., & SCHOENFELD, N. The influence of ego-involvement on confidence. *J. abnorm. soc. Psychol.*, 1941, **36**, 249–258.
68. KORSCH-ESCALONA, S. The effect of success and failure upon the level of aspiration and behavior in manic-depressive psychoses. *Univ. Ia. Stud. Child Welf.*, 1940, **16**, 199–302.
69. KULP, D. H. II. Prestige, as measured by single-experience changes and their permanency. *J. educ. Res.*, 1934, **27**, 663–672.
70. LANDAHL, H. D. A contribution to the mathematical biophysics of psychophysical discrimination. *Psychometrika*, 1938, **3**, 107–125.
71. LANDAHL, H. D. A contribution to the mathematical biophysics of psychophysical discrimination. III. *Bull. Math. Biophysics*, 1940, **2**, 73–87.
72. LANDAHL, H. D. Discrimination between temporally separated stimuli. *Bull. math. Biophysics*, 1940, **2**, 37–47.
73. LANIER, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, **11**, 199–217.
74. LEWIS, H. B. Studies in the principles of judgments and attitudes: IV. The operation of "prestige suggestion." *J. soc. Psychol.*, 1941, **14**, 229–256.
75. LONG, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. *Arch. Psychol.*, 1937, No. 209.
76. LORGE, I., & CURTIS, C. C. Prestige-suggestion and attitudes. *J. soc. Psychol.*, 1936, **7**, 386–402.
77. LUND, F. H. The psychology of belief. *J. abnorm. soc. Psychol.*, 1925, **20**, 180–190.
78. LURIE, W. Measurement of prestige and prestige-suggestibility. *J. soc. Psychol.*, 1938, **9**, 219–225.
79. MARPLE, C. H. The comparative susceptibility of three age groups to the suggestion of group versus expert opinion. *J. soc. Psychol.*, 1933, **4**, 176–186.
80. MCGARVEY, H. R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, No. 281.
81. MCGREGOR, D. M. The major determinants of the prediction of social events. *J. abnorm. soc. Psychol.*, 1938, **33**, 179–204.
82. MILLER, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1939, **52**, 562–578.
83. MOORE, H. T. The comparative influence of majority and expert opinion. *Amer. J. Psychol.*, 1921, **32**, 16–20.
84. MOSIER, C. I. A psychometric study of meaning. *J. soc. Psychol.* 1941, **13**, 123–140.
85. MURPHY, G., MURPHY, L. B., & NEWCOMB, T. M. *Experimental Social Psychology.* New York: Harper, 1937.

86. OSGOOD, C. E. Ease of individual judgment-processes in relation to polarization of attitudes in the culture. *J. soc. Psychol.*, 1941, **14**, 403–418.
87. OSGOOD, C. E., & STAGNER, R. Analysis of a prestige frame of reference by a gradient technique. *J. appl. Psychol.*, 1941, **25**, 275–290.
88. PAYNE, B., & DAVIS, R. C. The role of muscular tension in the comparison of lifted weights. *J. exp. Psychol.*, 1940, **27**, 227–242.
89. PINTNER, R., & FORLANO, G. The influence of attitude upon scaling of attitude items. *J. soc. Psychol.*, 1937, **8**, 39–45.
90. REMMERS, H. H., CUTLER, N., & JONES, P. Waking suggestibility in children: General or specific? *J. genet. Psychol.*, 1940, **56**, 87–93.
91. ROGERS, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.
92. SAADI, M., & FARNSWORTH, P. R. Degrees of acceptance of dogmatic statements and preferences for their supposed makers. *J. abnorm. soc. Psychol.*, 1934, **29**, 143–150.
93. SAFFIR, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 170–198.
94. SCHLOSBERG, H. A scale for the judgment of facial expressions. *J. exp. Psychol.*, 1941, **29**, 497–510.
95. SELLS, S. B. The atmosphere effect. *Arch. Psychol.*, 1936, No. 200.
96. SEWARD, G. Recognition time as a measure of confidence. *Arch. Psychol.*, 1928, No. 99.
97. SINGER, W. B., & YOUNG, P. T. Studies in affective reaction. III. The specificity of affective reactions. *J. gen. Psychol.*, 1941, **24**, 327–341.
98. STRONG, E. Effect of length of series on recognition memory. *Psychol. Rev.*, 1912, **19**, 447–462.
99. SYMONDS, P. M. Notes on rating. *J. appl. Psychol.*, 1925, **9**, 188–195.
100. SYMONDS, P. M. *Diagnosing Personality and Conduct.* New York: Century, 1931.
101. THORNDIKE, E. L. Fundamental theorems in judging men. *J. appl. Psychol.*, 1918, **2**, 67–76.
102. THORNDIKE, E. L. A constant error in psychological ratings. *J. appl. Psychol.*, 1920, **4**, 25–29.
103. THORNDIKE, E. L. Primitive forms of belief and knowledge. *Psychol. Rev.*, 1934, **41**, 403–411.
104. THORNTON, G. R. The effect upon judgments of personality traits of varying a single factor in a photograph. *J. soc. Psychol.*, 1943, **18**, 127–148.
105. THORNTON, G. R. The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *J. appl. Psychol.*, 1944, **28**, 203–207.
106. TROW, W. C. The psychology of confidence. *Arch. Psychol.*, 1923, No. 67.
107. TROW, W. C. Trait consistency and speed of decision. *Sch. & Soc.*, 1925, **21**, 538–542.
108. VOLKMANN, J. The relation of time of judgment to certainty of judgment. *Psychol Bull.*, 1934, **31**, 672–673.
109. VOLKMANN, J. Anchoring of absolute scales. *Psychol. Bull.*, 1936, **33**, 742–743.
110. VOLKMANN, J., HUNT, W. A., & MCGOURTY, M. Variability of judgment as a function of stimulus-density. *Amer. J. Psychol.*, 1940, **53**, 277–284.
111. WEBER, P. Judgment today. *Psychol. Rev.*, 1937, **44**, 264–266.
112. WEVER, E. G., & ZENER, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35**, 466–493.
113. WILLIAMS, A. Perception of subliminal visual stimuli. *J. Psychol.*, 1938, **5**, 187–199.
114. WOODROW, H. Weight discrimination with a varying standard. *Amer. J. Psychol.*, 1933, **45**, 391–416.
115. WOODWORTH, R. S. *Experimental Psychology.* New York: Henry Holt, 1938.
116. YOUNG, P. T. *Emotion in Man and Animal.* New York: Wiley, 1943.

PSYCHOLOGY AND THE WAR

Edited by

DONALD G. MARQUIS

CONTENTS

THE OFFICE OF PSYCHOLOGICAL PERSONNEL

Report for the Year 1944

DONALD G. MARQUIS

At the end of its third year of operation, the Office of Psychological Personnel has reached a certain degree of stability in its functions in spite of the rapidly changing conditions of wartime. The Office was created by recommendation of the Emergency Committee in Psychology in February 1942 to promote the maximum effective utilization of psychology and psychologists in the war effort. It has operated under the administrative supervision of the Division of Anthropology and Psychology of the National Research Council which also provided office space and facilities in its building at 2101 Constitution Avenue, Washington 25, D. C.

It is expected that the Office of Psychological Personnel will cease to exist at the end of 1945, when its activities will be assumed by the newly created Office of the Executive Secretary of the reorganized American Psychological Association. Looking ahead to this shift, the Office of Psychological Personnel has accepted an increasing participation in the professional activities of psychology and has cooperated in the work of several of the active committees of the national societies. The Director of the Office of Psychological Personnel has joined with the Secretary and Business Manager of the APA and the Executive Secretary of the AAAP in formulating a statement of possible functions of the Office of the Executive Secretary. This statement will be presented to the Council of Representatives of the APA at the September meeting and will be available for the use of the Executive Secretary when he takes office.

Psychologists in the Armed Services

During the course of the year there has been a noticeable shift in the nature of psychological work in the Army and Navy. The original demand for job analysis and for personnel selection and classification procedures has been satisfied in large part, and many psychologists have been transferred to duties in the analysis and design of training procedures and in clinical psychology. New programs have developed in the rehabilitation of military prisoners, in the special training of the inept, in the evaluation of combat effectiveness, in special researches in the theaters of operation, in the redistribution of personnel returned from combat duty, in the hospital services for disabled, and in the discharge (separation) counseling.

Relatively few psychologists have been inducted through Selective Service during the year. The available supply had been pretty well exhausted and the new class of 18 year olds does not, of course, include any psychologists. The urgent demand for young combat troops in the spring of the year resulted in the cancellation of occupational deferment for most psychologists under 26. The Office of Psychological Personnel continued to help new inductees obtain classification for psychological duties, and to assist others in obtaining commissions.

Direct commissions for enlisted psychologists with clinical training and experience became available during the latter half of the year in a developing program of hospital service jointly administered by the Surgeon General and the Adjutant General. The OPP responded to a request for information about all qualified psychologists, and furnished similar recommendations in connection with several other less extensive opportunities for direct commissions.

In spite of the corrective reassignments and increased quotas for commissions, complete utilization of professional psychologists in the Armed Forces still has not been realized. Approximately one tenth of those with Ph.D. training are not officers, and many with somewhat less training are not assigned to psychological duty. The high priority of assignment to Ground Forces this year has meant that any psychologists who for one reason or another found themselves in Infantry or Artillery were out of reach of any measures to reclaim them for research or psychological duties in other branches of the Army. These problems, and others concerned with the role of psychologists in the military organization, were the subject of study and recommendation by a series of conferences of military psychologists called by Dr. Robert M. Yerkes at the National Research Council.

Psychologists in Federal Agencies

The Office of Psychological Personnel has continued to consult with

the Government departments and war agencies needing psychological service, and to assist in locating personnel for their programs. A few of the sections employing psychologists have been curtailed; more have found that the limited personnel available has served to restrict their expansion.

The Civil Service Commission, looking ahead to the enormous employment task after the war, has organized a Test Development Section, with Thomas Bransford as Chief. Approximately 15 psychologists have been employed to date, and it is expected that others will be added to the staff as the work develops.

The number of psychologists engaged in various projects in the Office of Strategic Services has increased almost twofold during the year. Approximately half of them are in the Armed Forces. The Office of the Chief Engineer, War Department, has requested lists of qualified industrial psychologists throughout the country who would be available for consultation on regional personnel problems.

The progress of the war has resulted in a reduction in the work of the Surveys Division of the Office of Civilian Requirements, War Production Board, and Helen Peak has replaced E. R. Hilgard as Chief. Dwight Chapman has transferred to the Census Bureau to develop new lines of work in special surveys and attitude studies for other Government agencies. The Office of War Information has reduced the work of its domestic branch, including the special services in which several psychologists were engaged. There are about half a dozen psychologists in the overseas branch, working in the Washington, New York, London and Paris offices. Leonard Doob has been made Chief of Policy Coordination as well as Chief of the Bureau of Overseas Intelligence. In the Division of Program Surveys, Department of Agriculture, Angus Campbell is acting as Chief in the absence of Rensis Likert, who is on an overseas research project for the War Department.

Approximately 140 psychologists are now employed in war projects associated with the Office of Scientific Research and Development. The National Research Council's Committee on Selection and Training of Aircraft Pilots employs 19 others on its various researches.

The Office of Vocational Rehabilitation, Federal Security Administration, in its expanding program has added several psychologists to the Washington staff in the Vocational and the Research Divisions, and a few others have already been engaged in the operating program at the State level.

The greatest potential demand for psychologists is in the Veterans Administration. Four psychologists are now engaged in the planning and supervising functions of the Vocational Rehabilitation Division of the central headquarters and others have taken positions in the voca-

tional advisement work of the regional offices in some ten states. There has been steady progress in the establishment of decentralized vocational counseling centers in the colleges and universities of the country. In August the Office of Psychological Personnel issued to department chairmen a 7-page newsletter describing the newly created program and discussing the opportunity for contract services by colleges. Copies of this newsletter may be obtained from the Office on request.

Job Referral Services

Requests from prospective employers were received this year in comparable frequency with the previous year. Table I presents an analysis of these requests and the outcome as determined by a follow-up inquiry.

TABLE I

Analysis of Job Requests
1944

Employer	*Total requests*	*For information or no placement made*	*Placement made*	*Placement from OPP referrals*
Colleges, universities	81	41	40	16 (17 individuals)
Industries, consulting organizations	20	12	8	3
Schools	6	2	4	1
Clinics, guidance centers	25	11	14	6
State hospitals, other institutions	19	15	4	3
Public personnel agencies	7	4	3	3 (4 individuals)
Federal war agencies	21	14	7	6
Other federal agencies	11	5	6	4 (10 individuals)
War research projects (OSRD, NDRC, CMR, NRC)	8	1	7	4 (24 individuals)
Totals	198	105	93	46 (74 individuals)

Because of the acute shortage of qualified and experienced psychologists no attempt has been made to promote an increased employment of psychologists in jobs not directly concerned with war activities. Anyone who is available to consider a change in employment, either temporary or permanent, is urged to register with the Office on forms which will be supplied on request. At the September meeting of the APA a special

desk for the OPP was set up to furnish information and to allow for conversation with prospective clients. The general employment situation is reflected in the fact that from the numerous interviews there developed fourteen requests for help in obtaining psychologists and only four available applicants.

Information Services

The Office of Psychological Personnel continues to receive many requests for various types of information about psychology and psychologists. Many useful suggestions have also been received and acted upon or referred to appropriate officers or committees for action.

Several individuals and agencies have requested information about vocational opportunities in psychology. The Office has cooperated with the Information and Education Division, War Department, in preparing such information for distribution to soldiers through the U. S. Armed Forces Institute, and has worked with the National Roster of Scientific and Specialized Personnel in the formulation of publications for distribution to U. S. Employment Service offices, high schools, Veterans Administration offices, etc.

It is apparent that more complete and accurate information about careers in psychology is needed. The OPP is cooperating with the Subcommittee on Occupational Standards, Emergency Committee in Psychology, and with the APA Committee on Graduate and Professional Training in the collection and analysis of descriptions of the work and qualifications in the various specialized branches of psychology. The results of this study, combined with the statistical data of the OPP employment survey of January 1944, will form the basis for a pamphlet for the vocational guidance of those who are considering psychology as a life-work.

The following reports have been issued by the Office of Psychological Personnel during the year:

Social psychologists in National War Agencies, *Psychol. Bull.*, 1944, *41*, 115–126.

Office of Psychological Personnel. Report for the Year 1943. *Psychol. Bull.*, 1944, *41*, 246–252.

The mobilization of psychologists for war service. *Psychol. Bull.*, 1944, *41*, 469–473.

By Robert R. Sears. Clinical psychology in the military services. *Psychol. Bull.*, 1944, *41*, 502–509.

Post-war reemployment prospects in psychology. *Psychol. Bull.*, 1944, *41*, 653–663.

Staff of the Office

The staff of the Office of Psychological Personnel during the year 1944 included the following persons:

Director (part time), Donald G. Marquis; *Assistant to the Director*, Jane D. Morgan; *Secretarial and clerical*, Ruth Aldridge, Paula Canter, Evelyn Lees, Rita Quigley, and Georgiana Stevens; *Consultants*, Robert M. Yerkes, Alice I. Bryan, Willard C. Olson, and Robert R. Sears.

Financial Statement—1944

Receipts

Balance on hand	$ 300.00
American Psychological Association	10,000.00
American Association for Applied Psychology	1,000.00
	$11,300.00

Expenses

Salaries (Director, Assistant, secretarial and clerical help)	$ 4,719.57
Office expense (postage, mimeographing, printing, telephone and telegraph, supplies)	654.17
Travel	1,030.09
Reprints	212.68
Surveys	1,300.00
	$ 7,916.51
Balance	3,383.49
Refund made to the American Psychological Association	$ 3,075.90
Refund made to the American Association for Applied Psychology	307.59

SPECIAL TRAINING PROGRAM FOR COUNSELORS OF WAR VETERANS

JOHN GRAY PEATMAN
The City College of New York

A special training program for psychologists and college personnel concerned with the vocational guidance and advisement of war veterans was conducted by The City College of New York in cooperation with the Veterans Administration for a four-week period during November and December, 1944. We shall briefly summarize the purpose of this special program, describe its content, and indicate some of the implications of the decentralized guidance program of the Veterans Administration.

Purpose of the Program

The need for such special training programs arises out of the fact that many colleges and universities throughout the United States have been or are being encouraged by the Veterans Administration to play an integral role in the vocational rehabilitation of disabled veterans and in the reintegration of the non-disabled veteran to civilian life. This development was recently summarized by Mr. H. V. Stirling, Director of Vocational Rehabilitation and Education Service of the Veterans Administration, as follows:

> In order that our disabled veterans may be brought into contact with the best qualified vocational counselors in the country a plan has been inaugurated . . . designed to effect close cooperation between our educational institutions and the Veterans Administration in providing further decentralization of vocational rehabilitation activities so that vocational counseling and induction into training may be accomplished more conveniently and efficiently at points near the homes of the disabled veterans (1).

Guidance centers have already been established in a considerable number of colleges and universities, and a total of two hundred or more is planned by the Veterans Administration in its implementation of Public Laws 16 and 346, passed by Congress to aid both the disabled and non-disabled veteran in his return to civilian affairs. Both of these laws call for a national program of veteran education and vocational training. Public Law 16, adopted in March, 1943, set up a vocational rehabilitation program for disabled veterans. Public Law 346, adopted in June 1944, and popularly known as the G.I. Bill of Rights, is for all qualified veterans. Under both of these laws, administered by General Frank T. Hines of the Veterans Administration, the Federal Government is charged with the responsibility not only of providing educational and vocational training to disabled and non-disabled veterans, but also has the responsibility of making vocational advisement available to each

veteran. Under Public Law 16 the disabled veteran is required to have the benefit of such advisement in order that, insofar as is possible, vocational handicap may be overcome; under Public Law 346, the veteran is entitled to guidance if he feels he may benefit by the advice of counselors with a wider background of training and experience than his own.

Millions of veterans, disabled and non-disabled, will be entitled to educational or vocational training over a period of from one to four years with tuition costs and subsistence allowances paid for by the Federal Government. Initially, they will be entitled to—they will need to have in a great number of the cases—the benefit of informed guidance. It is obvious that the success of this Guidance Program, set up on a widespread national scale that challenges the imagination, will depend upon the availability, in colleges and universities, of counselors who have not only the general psychological training requisite to vocational guidance, but who also will be informed of the details and many implications of the particular problems of this situation. The ultimate success of this vast program of vocational rehabilitation and education will depend to a large extent upon the character and competence of the guidance initially given by the counselor to the veteran prior to his actual entrance into a training program.

The program originally set up to provide the counseling services required by Public Law No. 16 was developed by the Vocational Rehabilitation and Education Service of the Veterans Administration in the Spring of 1943. This program was organized to include the procedures whereby guidance for disabled veterans was to be provided through the Rehabilitation Division of the regional offices of this agency. The personnel who carried on this advisement were, therefore, in the employ of the Veterans Administration, and the procedures prepared for use in interviewing, objective testing, supplying occupational information and making the final selection of the employment objective were developed by guidance specialists of the Vocational Rehabilitation and Educational Service of this federal agency. A series of standardized forms for collecting pertinent information and facts and for recording the evaluation of these facts were also prepared by them. A manual of Vocational Advisement to direct the procedures to be used in guidance was issued in January, 1944. This agency then brought to Washington for orientation training in the use of this Manual and the related forms two groups of psychologists and counselors selected to fill positions as Advisors in the regional offices of the various states.

As the man-power shortage increased and the advisement program was enlarged in scope by the passing of the so-called G.I. Bill of Rights in June 1944, the problem of obtaining highly trained personnel needed

for this work in the regional offices of the Veterans Administration became acute. This led to the plan of decentralizing the guidance program to colleges and universities. According to this plan such educational institutions as already had the required trained personnel for rendering counseling services of this nature were to be requested by Veterans Administration representatives in the regional offices of their state to submit an Offer Form for providing counseling services to veterans in case they were interested in aligning themselves with this program. Once these Offer Forms, stating in detail the facilities available for rendering such service, and the estimated per capita charge to the government for this service, were submitted by a college or university to the Veterans Administration, contracts between the two organizations could be effected. Such an arrangement offered the advantage that the colleges and universities were in a more favorable position than the Veterans Administration to avail themselves of properly qualified personnel. In addition the colleges and universities would be more closely accessible to the veteran, since the plans were designed to provide for several guidance centers in each state.

The first guidance center was established at The City College of New York in June, 1944, and has served as the pioneer advisement unit in the development of the college program. By September the counseling procedures of this first guidance center were well organized and functioning smoothly, but only after much had been learned. An advisement process that would be both systematic and at the same time yield a result vital and satisfying to each individual veteran had to be mastered. The psychologists in charge of the work were well-trained and long-experienced in college advisement and personnel services. However, they found themselves called upon to counsel and advise men and women whose education and job experience, whose interests and aptitudes, whose personalities and physical condition were as varied as our society produces. It promptly became necessary for them to become thoroughly cognizant of the details of the procedures as developed and required by the Veterans Administration, including a knowledge of the physical and psychological demands of all kinds of occupations, the character of the educational or training requirements of all kinds of jobs, and the probable post-war trends in work opportunities. A great deal had to be learned and it couldn't be accomplished in a day.

The purpose of the special four-week training program offered by City College was to share with psychologists of other colleges and universities the benefits of the experience obtained in its own advisement and guidance procedures as worked out in practice with the veterans and with the Veterans Administration. It was felt that such a program would help expedite the initiation of guidance centers in other universi-

ties, and that returning veterans in other parts of the country would therefore have competent counseling services under Public Laws 16 and 346 sooner than otherwise would be the case. Announcement of the program was made early in October, and on November 13, professors of psychology, directors of college personnel services, and college deans were on hand representing the Universities of Arkansas, Buffalo, Cincinnati, Dayton, DePauw, Georgia Tech., Howard, John Carroll, Miami, Missouri, North Carolina, Pittsburgh, South Carolina, William and Mary, and Youngstown, as well as St. Elizabeth's Hospital in Washington, D. C., the American Nurses Association, and neighboring colleges and universities in this area and in New England. All of the psychologists registered for the program had had considerable experience in vocational guidance or college personnel work. All were Ph.D.'s with the exception of two, who had their M.A. in Psychology and were completing the requirements for the Ph.D. In most cases the psychologists present were released by their institutions to attend the program in order that, on their return, they might immediately take over the operation of guidance centers being established at their institutions under contract with the Veterans Administration at Washington.

Nature of the Program

The special program was organized as follows:

1. A series of lectures and discussions during the first week* (2).
2. Seminars in occupational information for the advisement of veterans.
3. Seminars in psychological tests used in the advisement of veterans.
4. Seminars in the study of case records from The City College Guidance Center.
5. Clinical observation and participation in the actual procedures at The City College Guidance Center.
6. Round tables on Saturday mornings with all members of the group and City College and Veterans Administration personnel, for a discussion of the problems and questions arising out of the preceding week's work.†

* These lectures were published as "Proceedings of the Special Training Program in the Advisement of Veterans," by the City College of New York (2).

† The personnel for the program consisted of the following: Dr. Daniel F. Brophy, Director of The City College Personnel Bureau and Director of The City College Guidance Center; Dr. Ira D. Scott, Chief of the Advisement and Guidance Division, Veterans Administration, Washington, D. C.; Drs. Louis Long, Kathryn Maxfield, LaVange Richardson and Frank Shuttleworth of The City College Personnel Bureau and Guidance Center; Drs. Marion R. Bartlett, Lorenz A. Meyers, James H. Russell, Carlos E. Ward, and Mr. Helmar S. Peterson, Veterans Administration, Washington, D. C.; and Dr. John Gray Peatman, Associate Dean and Associate Professor of Psychology of The City College, Director of the Special Program and Editor of the *Proceedings*. Dr. Gardner Murphy, Chairman of the Department of Psychology, The City College, and Lt. Hugh M. Bell of the Adjutant General's Department, Washington, D. C., also gave lectures during the first week of the program.

The lectures during the first week of the program served as an introduction to the basic principles and philosophy, as well as legal requirements, of the advisement procedures. The seminars and observation and participation in the clinical procedures at the guidance center were conducted concurrently during the latter three weeks of the program. The psychologists in attendance were subdivided into four groups in order that each member would have full opportunity to participate in all aspects of the program. The general character and principles of the guidance procedures were presented in considerable detail by Dr. Ira Scott who originally formulated the entire guidance program for the Veterans Administration, and particular emphasis was given to the determination of a veteran's need for training, as required under Public Law 16, and to the use of the Advisement Forms, 1902 series, employed in working out an appropriate occupational objective with the veteran. The kinds of counseling required were differentiated. Counselors need to be prepared to give personal adjustment guidance as well as vocational and educational guidance. The testing aspects of the procedures were presented in detail. A standardized battery is considered entirely out of place in this guidance program. Rather, psychological tests of various kinds for the measurement of abilities, aptitudes, discovery of interests and insights into personality are selected for use only as indicated—that is, only as relevant to the particular case. The seminar on occupations began with the use of the Dictionary of Occupations, required in the advisement procedure; the physical and psychological demands, the educational and training requirements, the employment possibilities of all kinds of occupations were studied. Towards the end of the program, each psychologist had the opportunity to engage in the actual advisement procedures at the guidance center, counseling one or more veterans.

Some Implications of the Veterans Program of College Guidance Centers

One of the chief implications of the college guidance centers being established by the Veterans Administration is that they should enable the veteran to have a level or quality of guidance and advisement far superior to anything of this kind ever provided by any nation to its war veterans. The reorientation and integration of the returning servicemen to civilian life should therefore be more satisfactory than could possibly be the case without the availability of guidance provided by highly-trained psychologists in colleges and universities. The long-run social implications are self-evident.

A second major implication of this program lies in the fact that many colleges and universities that have not had adequate personnel services

will doubtless be able to maintain, after the Veterans Program has ended, such a center not only for their own institutions but for their communities as well. In this connection, it should be recognized that guidance services of the kind being developed for the returning serviceman are also needed throughout the nation for the civilian war worker, who, in many cases, will also have a real problem of reorientation and integration into our post-war society. Some community service centers have already been established with the problems of both the veteran and the civilian worker in mind.

Finally, the demand for highly-trained psychologists to staff the guidance centers to be established at colleges and universities by the Veterans Administration far exceeds the present available supply. Hundreds will be needed not only for these institutions but also for the regional offices of the Veterans Administration.* The responsibility for meeting this demand for trained personnel is a challenge to many graduate schools of psychology. The training of men and women for this work obviously needs to be considerably more than academic. It requires not only adequate preparation in the use of tests for all levels of adult ability, but also an expansion of psychological programs to include courses in counseling techniques and in the study and use of occupational information, as well as interne experience in interviewing and advisement of adults.

BIBLIOGRAPHY

1. STIRLING, H. V. Education's part in the program of services for veterans. *Education for Victory*, 1944, 3, No. 12, Dec. 20, p. 9.
2. PEATMAN, J. G. (Ed.) *Proceedings of the special training program in the advisement of veterans.* New York: The City College of New York, 1944. Pp. 120.

* It is to be observed that the staff of a College Guidance Center includes a vocational advisor and training officer in the employment of the Veterans Administration, as well as the members in the employ of the institution itself.

Note on The Article

"Veterans Administration Vocational Training Program: Processing Procedures Used by the College of the City of New York"*

The Veterans Administration wishes to make brief comment upon the article by Drs. Brophy and Long, recently published in the *Bulletin*.* Misunderstanding as to the nature of the counseling procedures formulated and applied by this government agency for administering its vocational guidance program might very possibly occur from the recurrent use of the term *screening* in the article. The following statements from the article illustrate this point:

> The Veterans Administration has set up a screening process of which all disabled veterans making claims under Public Law 16 must avail themselves. . . . The screening and advisement process set up by the Veterans Administration was in the beginning operated by personnel of Veterans Administration itself. . . . If a veteran elects immediate screening, an appointment is made for him at the screening center.

The term *screening* literally means *to pass through a coarse sieve* and something of this connotation is retained in the modern idiomatic use of the term in industry and in the military and naval organizations. *Screening* as used by these organizations refers to processes used in the selection of personnel for certain types of jobs. Standards for the abilities or aptitudes required for a particular job are set up, and a technique, such as a brief battery of tests, is used to indicate whether or not men possess those abilities and aptitudes.

The Veterans Administration's counseling procedures are developed and applied from an entirely different standpoint. The view taken under this program is not that the object is to screen veterans to find those who will fit into one of a relatively few jobs. *All* veterans who come for guidance must be individually counseled in order to explore each veteran's abilities, aptitudes, and interests, with a view to finding which of all the vocational fields in the entire gamut of our social structure contains occupations suitable to his disability, abilities, etc., and to select one of these which seems most suitable to the veteran's needs, capacities and interests. Counseling and guidance techniques approach the problem from the viewpoint of the welfare of the individual person, whereas *screening* techniques are concerned with the job to be done. The two techniques are not always mutually exclusive, but differ widely in purpose, as *screening* has no place whatever in the techniques applicable to the counseling of veterans.

The article also stated that the Guidance Center at The College of the City of New York was to be operated as a *pilot* unit where "difficulties in the existing program could be discovered and worked out." While it is quite true that any program of this size has to be worked out empirically and perfected over a period of time, the fundamental groundwork had been laid by the Veterans Administration about a year prior to the making of a contract with The College of the City of New York for furnishing these services, and a very considerable number of veterans had already been counseled according to the procedures previously established. Forms for collecting the pertinent information and for logically evaluating it had been developed and a Manual of Vocational Advisement which explained the different phases of the work had been distributed to the Veterans Administration Regional Offices throughout the country. The manner in which the College of the City of New York did serve in a pioneer capacity was that of assisting in the working out of interrelationships between an educational institution and this government agency, rather than in *modifying procedures and correcting errors* as stated in the article.

* Brophy, D. F., & Long, L. Veterans Administration Vocational Training Program: Processing Procedures Used by C. C. of N. Y.: *Psychol. Bull.*, 1944, 41, 795–802.

THE DIFFERENTIAL VALIDITY AND DIFFICULTY OF SUBTESTS OF THE WECHSLER MENTAL ABILITY SCALE*

WILLIAM D. ALTUS, 1ST LT., A.G.D.†

Senior Personnel Consultant
Ninth Service Command Special Training Center
Camp McQuaide, California

The objectives of an Army Special Training Center have been previously described by Bell and Altus (1). Briefly, the main function of this and other like centers is to bring the trainee, within a specified time, to a degree of literacy in reading, writing and arithmetic, which is roughly analogous to that of the fourth-grade level in the public schools. If the trainee fails to reach this level within the prescribed time limit and has no occupational specialty useful to the Army, he is discharged as inapt. Although the percentage of men thus discharged varies somewhat from month to month, in the main it is relatively constant at this Center.

The dichotomous disposition of most of the trainees (a few are discharged for reasons of physical disability) permits a statistical analysis of the validity of the various measures used by the Personnel Consultants in sectioning, counselling and prediction. Six of the tests administered to each trainee on his arrival at the Center are of the personal interview type; two are group tests.

The Wechsler Mental Ability Scale, Form B, is one of the personally administered tests. The total scale consists of seven verbal and nine performance subtests. In order to equate the contribution of each subtest in the final weighting, the raw scores of each test are converted into derived scores with equivalent mean and standard deviation. Owing to the number of trainees arriving weekly and the paucity of trained interviewers, only a limited portion of the total Army Wechsler Scale is administered. Through a gradual process, the following procedure was evolved: All trainees who understand English well enough to be tested in that language are subjected to the Information, Comprehension, Similarities and Arithmetic subtests of the verbal portion of the Wechsler; those who are non-English or nearly so are given the Digit Symbol, Series Completion, Mazes and Block Designs subtests of the performance scale. This number of subtests is not sufficient to allow full confidence in the index of brightness derived therefrom, but it has proved, along with the other measures employed, of worth in serving the purpose for which it was intended.

The intercorrelations among the eight subtests of the Army Wechsler

* The opinions expressed in this article are those of the author and are not to be construed as reflecting the official attitude of the Army of the United States.

† 2d Lt. Ephraim Yohannan, 2d Lt. Jerry H. Clark and T/4 Edmund A. Ellis are to be credited with carrying out most of the statistical calculations in this article. Credit is also due T/4 Roy C. Burge and T/Sgt Clarence A. Mahler for aid in tabulating data.

administered at this Center are presented in Table I. The performance subtests generally have a considerably higher correlation with each other than they do with the verbal subtests; the correlations of the verbal subtests with each other are also lower than are the intercorrelations of the performance tests. The Digit Symbol and Series Completion have the highest intercorrelation among the performance tests with an *r* of .660; the highest comparable *r* among the verbal subtests is .462 between Similarities and Information.

TABLE I

INTERCORRELATIONS AMONG THE EIGHT SUBTESTS OF THE WECHSLER MENTAL ABILITY SCALE, FORM B.

Subtest	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. Information		.214	.447	.462	.288	.332	.295*	.111
2. Arithmetic	.214		.383	.231	.383	.339	.026*	.314
3. Comprehension	.447	.383		.356	.191	.264	.346*	.189
4. Similarities	.462	.231	.356		.193	.348	.447*	.106
5. Digit Symbol	.288	.383	.191	.193		.660	.570	.593
6. Series Completion	.332	.339	.264	.348	.660		.446	.456
7. Mazes	.295*	.026*	.346*	.447*	.570	.446		.531
8. Block Designs	.111	.314	.189	.106	.593	.456	.531	

* These *r*'s are based on an N of 52; all other *r*'s in the table have an N of 100.

One should be admonished that *r*'s in Table I do not represent the true interrelationship among these several subtests for at least two reasons. One is that the range of ability in an Army Special Training Center is especially narrow, almost all of it lying at the lower end of the distribution; this restriction causes the intercorrelations to be considerably attenuated. The other distorting factor is due to the fact that there are two groups represented in the matrix: the intercorrelations of the performance tests are derived from a non-English group; the correlations of the verbal subtests with each other and with the performance tests are derived from an English-speaking group. The interrelationship of the verbal and performance subtests for the non-English group was, of course, impossible to obtain.

All of the measures used by the Consultants' Section have been or are being validated against the criterion of graduation as opposed to discharge for inaptness. Admittedly, the criterion is only a partial one; it has, however, the merits of being definite and unbiassed, the promotion tests and the final examinations being of an objective type. Owing to the bifurcate nature of the criterion, bi-serial correlations and critical ratios have been the means employed to assess the relative validity of the Wechsler Scale and the other measures employed. The index of validity for each of the eight Wechsler subtests in use at this Center has been calculated. Since the calculations were performed with derived scores, it is possible to compare the subtests directly in terms of their difficulty.

I. The Verbal Subtests

TABLE II

Validity and Reliability of the Arithmetic, Information, Comprehension and Similarities Subtests.

Measure	*Arithmetic*	*Information*	*Comprehension*	*Similarities*
r_{bis}	.467	.406	.360	.334
P.E., r_{bis}	.018	.018	.019	.020
Reliability, Test-Re-test*	.640	.806	.709	.670
P.E. of r	.036	.021	.029	.034
Mean of Graduates	5.67	5.48	4.83	5.93
Mean of Discharges	3.98	4.12	3.36	4.76
Combined Means	5.31	5.19	4.48	5.68
Diff./S.E. Diff. (Means of Graduates and Discharges)	14.70	13.33	10.89	10.73

* Mean number of days between tests, 52.05. The N for the r's concerning reliabiltiy is 121. The N's for the bi-serial correlations are very much larger.

In considering the validity coefficients in Table II, one must remember that the bi-serial r's suffer an attenuating effect, partly because the subtests are short and hence not too reliable, but mainly because the type of men sent to a Special Training Center are quite homogeneous in verbal aptitude, *i.e.*, in the lowest seven per cent, in terms of scores, of those who take the Army General Classification Test. It is apparent, then, that these subtests possess relatively high validity, as validity is defined in this article.

It may seem somewhat puzzling that the Arithmetic subtest has a higher validity, both in terms of bi-serial correlation and of critical ratios, than any of the other subtests in spite of its having the lowest test-retest reliability. The probable reason for the relative superiority of the Arithmetic test is that one of the two group examinations for determining graduation is an arithmetic test partly devoted to simple number processes and partly to problem solving. From the standpoint of administration, the Arithmetic subtest is the easiest and requires the least time. Conversely, the Comprehension subtest is relatively difficult to administer and takes longer than any of the other verbal subtests; unfortunately, the bi-serial r's do not indicate that validity is necessarily a function of difficulty of administration.

The ratios resulting from the division of the difference between the mean scores of those graduated and those discharged by the standard error of the difference are in perfect consonance, in terms of maghitude, with the bi-serial correlations in Table II. All of the critical ratios indicate statistically reliable differences, since they are well above the conventional three.

Table III presents data concerning the relative difficulty of the several verbal subtests for these soldiers of depressed verbal capacity

who are sent to a Special Training Center. It is permissible to compare the mean scores directly, as has been said before, because they are scores with the same mean and variability.

TABLE III

CRITICAL RATIOS* DERIVED FROM THE DIFFERENCES IN MEAN SCORES ON THE ARITHMETIC, INFORMATION, COMPREHENSION AND SIMILARITIES SUBTESTS.

Subtest	*Comprehension*	*Information*	*Arithmetic*	*Similarities*
Comprehension...	—	10.44	11.53	17.14
Information.....			1.85	7.90
Arithmetic.....				5.52
Mean Score.....	4.48	5.19	5.31	5.68

* Diff./S.E. Diff.

All the ratios in Table III are statistically significant, excepting the one between Information and Arithmetic; even this ratio indicates that there are 97 chances out of 100 that the difference in mean score between these two variables is significant. The Comprehension subtest is the most difficult for the type of soldier assigned to this Center. It might be argued that the type of question found in the Comprehension subtest demands the greatest amount of abstraction, and for that reason is the most difficult. Inferentially, however, the information subtest requires the least in the way of abstracting ability and yet is more difficult than either Arithmetic or Similarities. Logically, the Comprehension and Similarities subtests are much alike in that both make demands on the ability to educe relationships; nevertheless, they are far apart in their difficulty value, the critical ratio of the difference of their means being 17.14. The author of this article has no reason he considers defensible for the differences in difficulty among the subtests. It can only be said that real differences do exist.

Data relative to the differences in degree of association of the verbal subtests with the criterion, graduation versus discharge, are presented in Table IV. The intercorrelation among the several test variables was taken into consideration in computing the significance of the differences among the *r*'s, in accordance with formula 108 of Peters and Van Voorhis (2). Only two of the critical ratios in Table IV reach the level of statistical significance, both of which include Arithmetic. There is, however, a reasonable degree of certainty (19 chances out of 20) that the Arithmetic subtest is more valid than the Information test for use in Centers such as this. There is also a fair probability that the Information test is more valid than either Comprehension or Similarities.

Comprehension appears to be a better test than Similarities; at least there are about four chances in five that it would have a consistently higher degree of association value.

TABLE IV

RATIOS DERIVED FROM DIVIDING THE DIFFERENCE BETWEEN THE BI-SERIAL CORRELATIONS OF THE FOUR VERBAL SUBTESTS, BY THE PROBABLE ERROR OF THE DIFFERENCE.

Subtests Compared	$\frac{D}{P.E._D}$	*Chances in 100 of a True Difference*
Arithmetic—Information	2.54	95
Arithmetic—Comprehension	4.86	100
Arithmetic—Similarities	5.54	100
Information—Comprehension	2.19	93
Information—Similarities	3.60	99
Comprehension—Similarities	1.13	78

II. THE PERFORMANCE SUBTESTS

It has been mentioned that those trainees who are non-English or who know too little English to be tested on the verbal section are given the four performance subtests, Digit Symbol, Series Completion, Mazes and Block Designs. Table V presents data on these performance tests which are comparable except for reliability coefficients with those given in Table II for the four verbal subtests.

TABLE V

VALIDITY OF THE DIGIT SYMBOL, SERIES COMPLETION, MAZES AND BLOCK DESIGNS SUBTESTS.

Measure	*Digit Symbol*	*Series Comple-tion*	*Mazes*	*Block Designs*
r_{bis}	.585	.464	.448	.379
P.E. of r_{bis}	.033	.038	.038	.040
Mean of Graduates	6.32	7.22	9.07	7.54
Mean of Discharges	3.72	4.82	6.36	5.65
Combined Mean	4.38	5.44	7.05	6.14
Diff./S.E. Diff. (Means of Discharges and Graduates)	9.49	7.16	7.77	6.38

The most noteworthy fact to be gleaned from Table V is the relatively striking validity of the Digit Symbol subtest in comparison with the others. The mean of those who graduated is 2.60 points higher in standard score than is the mean of those who were discharged. This figure is almost exactly synonymous with the sigma of the combined distribution of discharges and graduates, 2.62. The bi-serial correlation of .585 is higher by .118 than is the highest bi-serial (Table II) for the verbal tests. There are 98 chances out of 100 that this is a true difference; if the correction for the linear relationship between the two variables is applied, there are 99 chances out of 100.

The second most valid subtest is that of Series Completion. The Maze test is not far behind the Series Completion test in bi-serial relationship. Least valid of the four performance subtests is the Block Designs test.

Certain things about the non-English group should be pointed out before an attempt is made to rationalize the good showing of the Digit Symbol test. The non-English trainees are often literate in their native tongue, though they resemble the English-speaking trainees in not being literate in English as the Army defines literacy. The task of the non-English trainee is much harder than is that of the soldier who comes to the Center speaking English. The trainee of Mexican descent or nationality, the Navajo Indian, the Chinese—many of whom come to the Center almost devoid of English—must learn to read and write it to an approximate fourth-grade level and be able to understand the tongue well enough to carry out simple orders. That is a sizable requirement when one considers that only twelve weeks are allowed for the task.

The high validity of the Digit Symbol subtest may result from the fact that it is in miniature the learning of a new set of symbols somewhat analogous to learning a new language. The digits one to nine are presented on paper with, say, an *x* under one digit, an *O* under another, a straight line with a dot over it under another, and so on. After a brief demonstration, the subject writes the correct symbols under as many of the 93 digits which follow as he can in the two minutes' time allowed for the test. Inferentially, this type of activity is similar to learning a new combination of letters for a familiar object, as an American student of Spanish does when he learns to know *house* under the new symbol, *casa*.

The logical deduction of the preceding paragraph may, of course be factitious, for it is possible that the Digit Symbol test may be saturated with Spearman's G factor or with Thurstone's V or W factors. There are certain data in the third section which militate against such an interpretation, however.

Table VI presents the probabilities of true differences among the performance tests in their association with the criterion. It is probable that all of the critical ratios involving the Digit Symbol bi-serial are significant, even though two of them do not quite reach the mandatory four. It appears likely that the Block Designs subtest has a significantly

TABLE VI

RATIOS DERIVED FROM DIVIDING THE DIFFERENCES BETWEEN THE BI-SERIAL CORRELATIONS OF THE FOUR PERFORMANCE SUBTESTS BY THE PROBABLE ERROR OF THE DIFFERENCES.

Subtests Compared	$\frac{D}{P.E._D}$	*Chances in 100 of a True Difference*
Digit Symbol—Series Completion	3.78	99
Digit Symbol—Mazes	3.81	99
Digit Symbol—Block Designs	5.57	100
Series Completion—Mazes	.38	61
Series Completion—Block Designs	1.93	91
Mazes—Block Designs	1.73	88

lower validity than any of the other tests since the differences in association value are consistent and all quite high. There seems little to choose between the Series Completion and the Mazes.

The critical ratios in Table VII are sufficiently high to assure statistical significance for all of the differences between the means of the performance subtests. It is of interest that the order of difficulty of the four subtests corresponds roughly with their order of validity. It is improbable, however, that difficulty, *per se*, has anything to do with the validity, for this relationship did not hold for the verbal subtests (Table II).

TABLE VII

CRITICAL RATIOS* DERIVED FROM THE DIFFERENCES IN MEAN SCORES ON THE DIGIT SYMBOL, SERIES COMPLETION, MAZES AND BLOCK DESIGNS SUBTESTS.

Subtest	*Digit Symbol*	*Series Completion*	*Block Designs*	*Mazes*
Digit Symbol.....	—	5.35	9.21	12.48
Series Completion.....			3.41	7.09
Block Designs.....				4.14
Mean Score.....	4.38	5.44	6.14	7.05

* Difference in mean score divided by S.E. Diff.

There is a large difference in the mean scores of the Digit Symbol and the Block Designs subtests, 4.38 to 6.14. The critical ratio corresponding to this difference is 9.21. The difference could be ascribed, in part, to the fact that men unused to using a pencil, as is true of many of the trainees, are penalized on a paper and pencil test, such as the Digit Symbol, while they are not so penalized on the Block Designs test, where they only manipulate colored cubes. This argument loses its weight when the paper and pencil Maze test is considered, however; the mean score for this test is 7.05, making it even easier than the Block test.

III. VERBAL AND PERFORMANCE TESTS COMPARED

It might logically be reasoned that the performance tests of the Wechsler are more valid than the verbal ones because their order of association with the criterion is on the average somewhat higher in this study. Since different groups were used in validating the two types of tests, one English-speaking and one non-English, and since the task faced by the separate groups is not the same although the criterion for graduation is constant, it is unwise to assume too much about the comparability of the two sets of validating bi-serial correlations. Fortunately, it is not necessary to rely solely on inference in regard to the relative validities of the two types of tests, although the subsequent findings are not so definite as could be wished.

When the Special Training Center was first organized in September, 1943, by consolidating the five separate units in the Ninth Service Command, more subtests of the Army Wechsler were given than was possible later when the receipts of trainees became greater. For the first two or three months, all English-speaking trainees were given the verbal subtests previously discussed and in addition two or more of the performance subtests were administered. Thus for the English-speaking group, it is possible to compare the relative validities of the two types of tests directly, although the number of cases involved is small in comparison with the studies reported in Section I and II.

It will be noticed that with the exception of one bi-serial correlation, all of the *r*'s of the verbal tests in Table VIII have the same order of magnitude as they had in Table II. Where Similarities was third in Table II, it is first in this table. Since there are over sixteen times the number of cases in Table II as in Table VIII, the *r*'s from the former table are, obviously, the more reliable.

TABLE VIII

VALIDITY COEFFICIENTS AND CRITICAL RATIOS OF FOUR WECHSLER VERBAL SUBTESTS AND TWO PERFORMANCE TESTS.

Measure	*Similarities*	*Arithmetic*	*Information*	*Comprehension*	*Mazes*	*Block Designs*
r_{bis}	.528	.356	.277	.240	.137	.067
P.E. of r_{bis}	.063	.072	.075	.076	.136*	.079
Mean of Graduates	5.83	5.75	4.82	4.22	6.81	6.08
Mean of Discharges	3.71	4.50	3.97	3.23	6.00	5.77
Combined Mean	5.35	5.46	4.63	3.99	6.63	6.01
Diff./S.E. Diff. (Means of Discharges and Graduates)	4.82	2.84	2.50	1.98	.65	.57

* The N for the Maze test is smaller than for the others, for which a constant number was used.

The two performance subtests, Mazes and Blocks, have a degree of association that is only one P.E. removed from pure chance. While it is true that the probable error is in part a function of the number of cases involved, the number of cases required to make these two bi-serial correlations significant would be enormous. More important than the statistical significance of the performance *r*'s is their magnitude in comparison with those in Table V for the non-English group: The Maze test drops from .448 to .137; the Blocks from .379 to .067. It appears that these tests have little validity for the English-speaking trainees while having considerable for the non-English soldier. If this finding can be shown to hold for the other performance tests, it will be of considerable interest and significance.

TABLE IX

THE RELIABILITY OF THE DIFFERENCES IN MEAN SCORES OF FOUR VERBAL AND TWO PERFORMANCE SUBTESTS.

Subtests	$\frac{D}{S.E._D}$	*Chances in 100 of a True Difference*
Blocks—Information	5.02	100
Blocks—Arithmetic	1.91	97
Blocks—Comprehension	6.56	100
Blocks—Similarities	2.16	98
Blocks—Mazes	1.07	86
Mazes—Information	3.64	100
Mazes—Arithmetic	2.10	98
Mazes—Comprehension	4.66	100
Mazes—Similarities	2.26	99

Table IX shows that *for the English-speaking trainee* the two performance subtests are consistently, and in some instances significantly, easier than the verbal subtests. Tests can be equalized by the use of standard scores when the whole range of talent is measured; however, the equivalence of standard scores appears to break down when a large number of cases, falling in the lower extreme of the distribution, is tested.

TABLE X

VALIDITY COEFFICIENTS AND CRITICAL RATIOS OF FOUR WECHSLER VERBAL SUBTESTS AND TWO PERFORMANCE TESTS

Measure	*Arithmetic*	*Comprehension*	*Similarities*	*Information*	*Digit Symbol*	*Series Completion*
r_{bis}	.754	.603	.505	.497	.281	.191
P.E. of r_{bis}	.058	.082	.091	.092	.106	.109
Mean of Graduates	5.23	4.17	5.25	5.11	4.79	5.40
Mean of Discharges	2.00	1.63	3.31	3.25	3.82	4.56
Combined Mean	4.71	3.58	4.80	4.67	4.57	5.20
Diff./S.E. Diff. (Means of Graduates and Discharges)	5.37	4.80	3.43	3.65	1.83	1.02

One may note in Table X that the validity coefficients of the verbal subtests are generally higher than they were in the previous tables. Excepting, however, the interchange of position of Comprehension and Information, their order of magnitude is the same as it was in Table II. The number of cases involved in Table X is less than four percent of the number in Table II; for that reason it is surprising that the validity of the subtests remains as constant as it does.

The Digit Symbol and Series Completion subtests show some validity and their order of association is somewhat higher than was true of the Mazes and Block Designs in Table VIII. The order of magnitude of the validity coefficients for the performance tests in Tables VIII and X is the same as it was for the non-English trainees in Table V. Where the coefficient for the Digit Symbol subtest was .585 (Table V, non-English trainees), in Table X it is .281; Series Completion drops from .464 to .191; Mazes, from .448 to .137; Block Designs from .379 to .067. This drop indicates that the performance tests, especially the Digit Symbol test, are neither tapping some generalized ability equally valid for all types of trainees nor some specialized factor such as Thurstone's V or W. By squaring the bi-serial *r*'s, one finds that the percentage of factors common to the Digit Symbol test and the criterion is approximately 34 for the non-English trainees; for the English-speaking group it is slightly less than eight. Similar drops will be noted for the other three performance tests. The validity of the performance subtests of the Wechsler is apparently governed by the degree to which the trainee in this Center is conversant with English.

The mean score of the Digit Symbol subtest shows this test (Table X) to be as difficult as the verbal ones. The Block and Maze tests appear, however, to be easier. While the mean score of the Digit Symbol is higher than the mean of the Comprehension test, it is lower than the means of the Information, Arithmetic and Similarities subtests. The mean standard score for the non-English trainees (Table V) on the Digit test is 4.38; for the English-speaking soldiers the score is 4.57, slightly higher but not significantly so.

The mean of the Series Completion subtest is higher than those of the other subtests in Table X. The only difference between the mean of the Series Completion and those of the other tests which is significant is with Comprehension, the critical ratio being 4.10. The non-English mean on the Series Completion (Table V) is 5.44, quite similar to the mean of 5.20 for the English-speaking trainees.

IV. Discussion

The most striking finding of this study is the surprisingly high validity of the eight subtests of the Wechsler Mental Ability Scale, Form B, when validity is defined as association with the criterion of graduation versus discharge for trainees in an Army Special Training Center.

The correlations may not appear on the surface to be high in the usual sense but the attenuation introduced by the shortness of the subtests and the very restricted range of aptitude possessed by the trainees gives point to the fact that the bi-serials markedly understate the true validity of the tests.

The very good showing of the Arithmetic among the verbal subtests is somewhat spurious, owing, as has been mentioned before, to the overlapping of the test variable with one of the two tests used as a criterion for graduation. Its ease of administration and its power of discrimination mark this test as the most valuable verbal subtest in the Army Wechsler for use in a Special Training Center.

In the mimeographed directions for the Army Wechsler (3), the Similarities subtest is rated second in validity. This study has shown it, however, to be the least valid (Table II) of the four verbal subtests compared. On the other hand, the Information test, listed as fourth in value by the mimeographed set of directions, shows an order of association second only to the Arithmetic subtest.

The Block Designs subtest, rated first among the performance subtests in the directions for administration, comes out a poor fourth in this study (Table V). The Digit Symbol and Series Completion tests place first and second (Table V); the suggested order of validity in the directions was second and third, respectively.

The Digit Symbol test is significantly more difficult than any of the other three performance subtests for trainees in this Center. There is a high degree of certitude that it is more valid than any of the other performance tests. It is also the speediest of the tests from the standpoint of administration.

There was no evidence (Section III) for a high saturation of a generalized factor, such as Spearman's G, or of a restricted verbal factor, such as Thurstone's V or W, in the Digit Symbol subtest. Its high validity when used with non-English trainees is probably due to the similarity of learning new symbols for old and deeply ingrained associations, common to the test situation and the learning of a new language. Not mentioned in the preceding sections is the fact that the Symbol test had a higher order of validity for non-English trainees of Mexican ancestry or nationality than did a test of their literacy in their native tongue, Spanish. One wonders whether research would demonstrate a like degree of validity for the Symbol test in measuring aptitude for foreign languages among high school and college students. Certainly the secondary schools and colleges would furnish a wider range of aptitude, the narrowing of which attenuates the correlations in this study.

It is of marked interest that the performance tests in this study correlate well with the criterion when employed with non-English soldiers but that they lose their value, compared with verbal tests, when the subjects are English-speaking. Even the Symbol test had a low validity in comparison with the verbal subtests when applied to trainees who speak English; it did, however, remain the most valid of the performance tests.

There are marked differences in the difficulty value of the Wechsler subtests, despite the fact that they are couched in terms of standard scores. The order of difficulty for the trainees at this Center is, from most to least difficult of the verbal tests, Comprehension, Information, Arithmetic and Similarities; the order for the performance subtests is Digit Symbol, Series Completion, Block Designs and Mazes. One may infer that standard scores of equal value may be regarded as equal *only when the total range of ability measured by the test is represented in the mean scores of the groups compared.* Some unpredictable shifts in difficulty value may occur when the subjects tested are at the lower extreme of the distribution.

The data presented in this study are derived from only one of the Army Special Training Centers. In spite of this parochial limitation, it is believed that the number of cases involved, especially in Sections I and II, is of sufficient magnitude to assure that the results obtained have more than a modest degree of validity and reliability for Centers of this type.

BIBLIOGRAPHY

1. Bell, H. M., & Altus, W. D. The Work of Psychologists in the Ninth Service Command Special Training Center. *Psychol. Bull.*, 1944, **41**, 187–191.
2. Peters, C. C. & Van Voorhis, W. R. *Statistical Procedures and Their Mathematical Bases.* New York: McGraw-Hill, 1940.
3. *Manual of Directions for the Administration of the Wechsler Mental Ability Scale, Form B.* (A manual published for Army use: no publication date or other data given.)

BOOK REVIEWS

MARTIN, LILLIEN J. *A handbook for old age counsellors.* San Francisco: Geertz Printing Co., 1944. Pp. v+84.

This posthumously published little volume, written in her 91st year, is the final statement by a remarkable woman regarding the work which made her famous.

After a foreword by Clare deGruchy there is a very brief historical introduction on "the development of social welfare work with old people." Part I then deals with "methods for leading the old person to develop greater insight into his own problems." Part II deals with "re-education of the client for active participation in community and industrial life." Part III concludes with "the continued self training of the counselor." There are included such simple practical aids as a suggested chart of expenditures for old people of various incomes, a list of "slogans and poems used in the old age counseling center," and "mental setting up exercises" for rehabilitation. A tabular summary showing the re-employment brought about for certain old persons and resulting gains both to the individual and the community, and a brief bibliography, end the volume.

As the above topics indicate, the volume's purpose is very briefly to give methods and points of view developed by Dr. Martin in her work with old people. This it seems to do admirably. Pervasive throughout is the author's realistic appreciation of the problems of old age, her commonsense and practical point of view in approaching them, and above all her magnificent courage. There are sundry illustrative anecdotes; style is straightforward with occasional phrases of a touching eloquence. Emphasis throughout is on helping the older person toward self-understanding, self-help, and usefulness. The volume should be on the list of every person interested in problems of the older ages, and may become a little classic in its field.

Various features of the program may be questioned: the testing program; the emphasis on self-help to a point that family and community resources for aiding the elderly are neglected almost to the point of positive rejection, a hopefulness regarding vocational usefulness that seems beyond the facts of a peacetime economy. But these last are perhaps correctives of common misconceptions in the opposite direction, and are characteristic of the author's sturdy courage.

SIDNEY L. PRESSEY.

Ohio State University.

KLUCKHOHN, CLYDE. *Navaho witchcraft.* Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University. Vol. XXII, No. 2, 1944.

Any study by Kluckhohn in Southwestern Indian ethnology should be informative, thought-provoking, and scientifically competent. The reader of this monograph will not be disappointed. If the critic feels disposed to offer any adverse criticism, as he can in considering the intepretative sections of the work, the author disarms him by anticipating most of the objections. He does this by stating clearly the limited objectives of the investigation and the assumptions upon which his interpretations rest. Moreover, he expressly disclaims anything more than a tentative, hypothetical status for his conclusions.

The monograph is concerned with "Navaho idea and action patterns con-

cerned with the influencing of events by supernatural techniques that are socially disapproved." The author deals separately with four varieties of witchcraft, approximately named Witchery, Sorcery, Wizardry, and Frenzy Witchcraft. All are evil-working activities, but they are distinguished one from another by differences in technique. These categories, by the way, were not invented by the author, but by the Indians themselves. The methods, the victims, the practitioners, and the motives are all described, and they make fascinating reading. One misses an adequate comparative analysis of the data, especially with reference to other southwestern peoples, but the author has specifically avoided such a large undertaking. There is just enough historical treatment to whet the reader's appetite for more. This point is particularly important, for the author's psychological interpretations are based upon the postulate of adjustment and adaptation. If this postulate is sound, an adequate historical analysis should reveal some correlation between changes in witchcraft beliefs and historical changes in social patterns. The author recognizes this fact, and does mention a few instances of such correlation.

The basis for the entire interpretative discussion is that cultural forms persist only if, in some sense, they contribute to the survival of the society or its members. In addition, they must help the individual to adjust, i.e., to facilitate responses which remove the motivation stimulating him. The actual analysis is done principally in psychological and psychiatric terms, always in a setting of Navaho social and economic living conditions. This last point is a saving one, for it makes the interpretation less purely speculative than it might otherwise have been. There is, however, a great deal of speculation. It is plausible, and its persuasiveness is reinforced by frequent references to the physical and economic circumstances of Navaho society, conditions which might well lead to witchcraft relief as an instrument of social regulation and control. In this connection, it should be noted that it is belief in witchcraft that is fairly widespread among the Navaho, not the practice of witchcraft. It has proved extremely difficult to establish the existence of witchcraft practice, although the author believes that some of its does persist.

The interpretative section is rich in suggestions which are based upon an intimate knowledge of Navaho culture. Being witched is, for instance, an excellent way for persons of low social status to win attention to themselves, thereby reestablishing their egos. Witchcraft belief and practices may also permit the expression of antagonisms which could not be expressed in other ways. Many factors in Navaho life tend to produce tensions and hostility, and the witch is a scapegoat that fits the pattern of Navaho thought. Witchcraft belief, practices, and accusations may be related to the mechanisms of identification and projection, and serve to relieve anxiety. As an instrument of social control witchcraft accusations may be important. For instance, it is a threat held over the heads of agitators and other social disturbers. The rich are always in danger of being accused of obtaining their wealth by witchcraft. This may tend to reduce glaring economic inequalities.

All of these hypotheses, of course, remain unverified, reasonable though they may be. The whole discussion is functional in scheme. It is easy enough to think of alternative functional interpretations, and of others that are not functional at all. A postulate different from Kluckhohn's basic one of adjustment and adaptation might lead to a very different line of explanation and interpretation. This different line would also stand in need of verification. Consistency and congruity are not enough. Nevertheless, this monograph is a noteworthy essay in the analysis of ethnological and sociological data in psychological

terms, even though the interpretation is almost altogether speculative. Social psychologists will do well to read this paper.

MATTHEW M. R. SCHNECK.

University of Arizona.

SLADEN, FRANK J. (Ed.) *Psychiatry and the war. A survey of the significance of psychiatry and its relation to disturbances in human behavior to help provide for the present war effort and for post war needs.* Springfield, Ill. & Baltimore: Charles C Thomas, 1943. Pp. xxii+505.

As a result of a grant from the McGregor Fund of Detroit a *conference on psychiatry* was held at the University of Michigan over a period of three days in the fall of 1942. The present volume is a record of the proceedings of this conference.

Some thirty odd specialists participated by reading papers at the conference or by discussing them. The range of topics was wide and the viewpoints diverse. As might be expected, much of the material introduced was neither new nor directly relevant to the central theme of maximizing psychiatric contributions to the war effort. The individual papers are so varied in style and content that unified appraisal of the volume as a whole does not lend itself to succinct and adequate formulation. As edited the book reveals a division of the work of the conference into five parts each of which merits a few words of descriptive comment by way of delineating the organization of the book.

Part I is devoted to the *Philosophy of Psychiatry* and consists of six essays. In the first one Adolf Meyer discusses the meaning and scope of psychiatry. The other five have to do with psychiatry's "relationship to psychological schools of thought" and to its significance in the practice of internal medicine, general surgery, pediatrics, and geriatrics.

Another group of six essays supplies the material for Part II which is concerned with *problems of research in psychiatry*. Psychologists are apt to be particularly interested in this portion of the book so that a slightly more detailed account of it might be justified. The initial essay by Cyrus C. Sturgis is a very general but nevertheless decidedly stimulating discussion of some phases of medical research. This is followed by C. Macfie Campbell's account of the "controversial in psychiatry." In this context Campbell introduces some pertinent observations with respect to schizophrenia. For him the controversy regarding this psychosis "has some of the aimless quality of a battle royal or melee. There is active discussion and criticism; but at the same time 'schizophrenia' remains merely a name, and one does not know to what the statements refer." He indicates that what is diagnosed as schizophrenia by the psychiatric staff of one institution might be differently diagnosed elsewhere. For example, in Burghölzli, famous because of its association with Bleuler, in the year 1940 of 578 first admissions 196 were placed in the schizophrenic category while only 2 were classified as manic-depressive. A ratio such as this is manifestly out of line with trends at other institutions. To read "reports of the results of various treatments," Campbell writes, "may seem interesting and may be accepted; but you are very anxious to know just what it is they have been treating. You have no idea." Because of this confused state of affairs he deems it impossible to evaluate the efficacy of the various techniques of shock treatment; since "it is extremely difficult to know what has actually been under treatment."

Campbell's paper is followed by Norman Cameron's discussion of the role of psychological research in psychiatry. He stresses a need for rapproache-

ment in saying, "I am convinced that psychology and psychiatry should never have parted company; they need each other badly." In addition, he points out the desirability of eliminating issues of hierarchical status in such cooperative research by suggesting that instead of regarding the psychologist as "merely the psychiatrist's *assistant*" the "two persons should be colleagues whose functions are regarded as of equal status" akin to those governing the professional teamwork of pathologists and surgeons. The need for teamwork is further suggested by Gildea's paper on the relevance of biochemical investigations for psychiatric progress. Incidentally, Cameron deplores the rise of *psychosomatic medicine*. For him it is "nothing but *mind-body* medicine" and "only new upholstery for a rickety piece of philosophical furniture; the old metaphysical structure is still there under new covering." However, about a dozen pages later one comes across a vigorous defense of "psychosomatic research in psychiatry" written by Franz Alexander. His justification for the retention of the concept Cameron spurns is indicated by the following passage:

> Psychosomatic research deals with such processes in which certain links in the causal chain of events lend themselves, at the present state of our knowledge, more readily to a study by psychological methods than by physiological methods since the detailed investigation of emotions as brain processes is not far enough advanced. My expectation is, however, that even when the physiological basis of psychological phenomena will be better known we will not be able to dispense with their psychological study. It is hardly conceivable that the different moves of two chess players can ever be more clearly understood in biochemical or neuro-physiological than in psychological and logical terms (127).

By implication support for Alexander's position is supplied by Nolan D. C. Lewis in his paper on "the future in psychiatry." He contends that such subjects as neuro-physiology and biochemistry are not basic to psychiatry. They are ancillary; for psychiatry itself is a basic science. His viewpoint is summed up in this sentence:

> It is a science of personal relationships, and while it is not a science of which every aspect can be subjected to laboratory experimentation to obtain the facts, we do have the facts of experience which are equally important and which we can handle in a scientific manner, substituting the experiment of the laboratory with the tool known as critical analysis (137–138).

As the foregoing excerpts indicate, the field of psychiatry is still plagued by conceptual and methodological problems whose metaphysical roots should be obvious to anyone familiar with the history of psychology. Parenthetically, it might be added that the entire volume reflects the divergent interpretations emerging from such roots.

To resume the task of sketching the volume's contents Part III must be considered. This is a potpourri of eight articles dealing with *psychiatry in the training, experience and education of the individual*. Although some of them are well executed summaries, the professional reader will discover little that is new to him as he peruses these articles. In less than 100 pages he is whisked through talks on psychiatry and education in general, on the part played by elementary schools, on the role of psychiatry in secondary schools and colleges, on similar roles in connection with courtship and marriage, in family life, in religion, in rural and urban community life and ending with William Healy's able discussion of the relationship of psychiatry to sociology and criminology.

Not until the middle of the book is reached does one come to grips with the subject of *psychiatry and the war*. The latter phrase is not only the title of the

book, but also the major caption for Part IV. Ten articles comprise this section. They include digests of the more or less well known ways in which psychiatric knowledge can enhance the efficiency of the army, the navy, the air force, and workers in civilian defense. There are also papers on national and international relationships, on postwar perspectives, on preventive psychiatry, on psychiatry in industry, and on morale and propaganda.

The fifth part of the book is made up of a review of the topics mentioned in the previous four parts with an account of the discussions of these topics by the numerous specialists who participated in the two symposia sponsored by the conference.

Considering the variegated nature of the material incorporated into these many articles the preparation of the index must have been a tremendous chore. The editor is to be congratulated for having executed this with exceptional diligence. Almost fifty pages are required for his splendidly detailed and usable index. This is all the more praiseworthy in view of the fact that volumes of this sort are so frequently published without any index at all. The care with which the volume was prepared for publication is also reflected in the relative paucity of typographical errors. The student of Greek might shudder to see *phylogenetic* spelled *philogenetic* (185). Descendants of Hans Gross, the criminologist, might also shudder to see their ancestor's name spelled *Huns Gross* (230). And no self-respecting group of psychiatrists would relish being called *phychiatrists* (288). Errors of this kind are hard to screen out in the first printing of a book and their occurrence is so rare in this volume that these few ought not to detract too markedly from one's appreciation of the competent craftsmanship shown by both editor and publisher in making these proceedings available to the public in attractive and readable form.

D. B. KLEIN.

University of Texas.

NOTES AND NEWS

WILLIAM THOMAS ROOT, JR., dean of the Graduate School of the University of Pittsburgh, succumbed to a heart attack, January 24, at the age of sixty-two years. Dr. Root had served as an instructor in psychology at Stanford University (1913–18), and at the University of Pittsburgh as professor of psychology since 1920, as head of the department since 1929, and as dean of the Graduate School since 1935.

SIDNEY CLARENCE GARRISON, president, the George Peabody College for Teachers (Nashville), succumbed to a heart attack, January 18, at the age of fifty-seven years. He was professor of educational psychology and director of instruction in the Senior College and Graduate School of George Peabody College until 1937, when he became president.

LYDIARD H. HORTON, consulting psychologist, Boston, and lecturer on biopsychology at the Boston University School of Medicine, died on January 19, at the age of sixty-five years.

CAROLINE BEAUMONT ZACHRY, since 1942 director, Bureau of Child Guidance, New York City Board of Education, died, February 22, at the age of fifty years. Dr. Zachry has served as an instructor (1922–23), Lincoln School, Teachers College, Columbia University; associate professor of psychology and head of the department of psychology and mental hygiene (1925–34), New Jersey Teachers College (Upper Montclair); director (1934–39), Research Commission on Secondary-School Curriculum and chairman of the study of adolescents, Progressive Education Association; and director (1943), Institute for the Study of Personality Development.

CLARENCE H. GRAHAM, professor of psychology at Brown University, has been appointed professor of psychology at Columbia University to take effect in September, 1945. He will be in charge of graduate work in experimental psychology.

At the Ohio University (Athens), J. R. GENTRY, GAIGE B. PAULSON, and T. C. SCOTT, have been promoted to professorships of psychology.

G. RICHARD WENDT, head of the department of psychology, Wesleyan University, will become professor of psychology and head of the department, University of Rochester (N. Y.), March 1, to succeed ELMER A. CULLER, who is resigning from the chairmanship because of ill health. Dr. Culler will, however, continue as professor of psychology and director of the hearing laboratory.

Since January 1, CHRISTIAN A. RUCKMICK has been assigned to the Operating Department of the Chicago and Northwestern Railway Company reporting directly to the vice-president of operations as supervisor of training. The scope of the training program covers the entire railroad and his services as psychologist are available to all departments of the company.

NORAH STEVENS, formerly psychologist with the YWCA, Toronto, was recently appointed to the staff of the Cottage School for Delinquent Girls, Sweetwater, Quebec.

LELAND H. STOTT, formerly assistant professor of home economics, University of Nebraska, has accepted a post at the Merrill Palmer School, Detroit.

BLANCHE CARRIER has been appointed instructor in psychology and sociology at the San Jose (Calif.) State College.

CHARLOTTE FEHLMAN has been appointed instructor in psychology, Adelphi College, Garden City, N. Y.

HUGBERT HAMILTON, associate professor of psychology at Temple Univer-

sity, has received a grant from the Temple University Research Fund for testing the field and growth hypothesis that body temperature affects maze learning.

LT. JOHN W. M. ROTHNEY recently received the following Commendation for Outstanding Services conferred by Brigadier General J. W. Barnett, U.S.A.:

First Lieutenant John W. M. Rothney, 0910881, AC, for outstanding service, 16 June 1944 to 21 December 1944. The splendid manner in which you inaugurated and expanded the educational program within this command exhibits outstanding leadership, ingenuity, and sustained effort. The coordination between the services, the assistance rendered the American Red Cross, and the competent management of all instructors in your classes have enabled limited materials to bring maximum results. Your untiring efforts and unselfish devotion to duty is worthy of the highest commendation.

The Rorschach Test course at Michael Reese Hospital is scheduled this year for the week of June 4–8, inclusive. The records to be demonstrated will be representative of the older adolescent and younger adult; with especial emphasis on persons discharged from the military services. DR. S. J. BECK, head of the psychology laboratory, will conduct the course. It meets twice daily, two hours each session. Interested persons may inquire of the Secretary, Department of Neuropsychiatry, at the Hospital, 29th Street and Ellis Avenue, Chicago 16.

A Syllabus for courses in *Psychology and the Problems of the Post-War World* is being prepared by a Committee of the Society for the Psychological Study of Social Issues and will be ready for use in Fall classes. The syllabus material will be integrated with the new SPSSI Yearbook on *Human Nature and Enduring Peace* published by arrangement with Reynal and Hitchcock and Houghton Mifflin, and will contain other reference material as well. Questions and suggestions for the guidance of the Committee should be addressed to one of the following: HORACE B. ENGLISH, The Ohio State University, Columbus, Ohio; DANIEL KATZ, Brooklyn College, Brooklyn, N. Y.; EUGENE HARTLEY, 3772 Earlham Street, San Diego, Calif.; CLYDE HART, 2517 Mozart Place, N. W., Washington, D. C.; HELEN PEAK, Meridian Hill Hotel, 16th and Euclid, Washington, D. C.

Correction. In the January *Notes and News* section, an item which stated that W. M. DANNER had been added to the staff of Oberlin College, should have read that Dr. Danner had been added to the staff of the University of California at Los Angeles as lecturer in psychology.

Plans for the APA Annual Meeting: Because of the regulations of the Office of Defense Transportation, the Council of Directors of the American Psychological Association has voted to postpone for the present its plans for the Annual Meeting. The regulations allow small groups to get together for the transaction of essential business and set up no limitations as long as the number meeting is less than fifty. The Council of Directors and the Program Committee stand ready to plan some type of program on short notice when conditions permit. If the ban on conventions is still in effect as the time for the Annual Meeting in September approaches, the present Council of Directors probably will arrange for a small meeting of outgoing and incoming officers in order to have a good discussion of important problems and to permit a smooth transition to the new plan of organization. It is anticipated that some of the business can be transacted, if necessary, under the *convention by mail* plan. WILLARD C. OLSON, *Secretary, American Psychological Association.*